



19 BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENT- UND  
MARKENAMT

12 Übersetzung der  
europäischen Patentschrift  
97 EP 0 778 582 B 1  
10 DE 690 33 862 T 2

61 Int. Cl.<sup>7</sup>:  
G 11 C 16/06

- 21 Deutsches Aktenzeichen: 690 33 862.7  
96 Europäisches Aktenzeichen: 96 114 368.2  
96 Europäischer Anmeldetag: 12. 4. 1990  
97 Erstveröffentlichung durch das EPA: 11. 6. 1997  
97 Veröffentlichungstag  
der Patenterteilung beim EPA: 14. 11. 2001  
47 Veröffentlichungstag im Patentblatt: 13. 6. 2002

DE 690 33 862 T 2

- |  |   |
|--|---|
| <p>30 Unionspriorität:<br/>337579 13. 04. 1989 US<br/>508273 11. 04. 1990 US</p> <p>73 Patentinhaber:<br/>SanDisk Corp., Santa Clara, Calif., US</p> <p>74 Vertreter:<br/>Hoffmann, E., Dipl.-Ing., Pat.-Anw., 82166<br/>Gräfelfing</p> <p>84 Benannte Vertragsstaaten:<br/>DE, FR, GB, IT, NL</p> | <p>72 Erfinder:<br/>Sanjay Mehrotra,, Milpitas, US; Eliyahou Harari,,<br/>Los Gatos, US</p> |
|--|---|

54 Effiziente parallele Programmierung von Datenblöcken in Multizustandsausführung

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

DE 690 33 862 T 2

EP 0 778 582

5

## Hintergrund der Erfindung

Diese Erfindung bezieht sich allgemein auf elektrisch löschbare, programmierbare Halbleiter-Festwertspeicher (EEPROM) und betrifft insbesondere Schaltungen und Techniken zum Lesen und  
10 Programmieren des Zustandes derselben.

EEPROM und elektrisch programmierbare Festwertspeicher (EPROM) werden typischerweise in digitalen Schaltungen für das nichtflüchtige Speichern von Daten oder Programmen benutzt. Sie können gelöscht werden, und es können neue Daten in ihre Speicherzellen eingegeben oder  
15 "programmiert" werden.

Ein EPROM arbeitet mit einem schwebenden (nicht angeschlossenen), leitfähigen Gate in einer Feldeffekttransistorkonstruktion, welches über einem Kanalbereich, aber von diesem isoliert, in einem Halbleitersubstrat zwischen einem Source- und einem Drainbereich angeordnet ist. Über  
20 dem Schweb-Gate, aber gleichfalls isoliert von demselben, ist ein Steuer-Gate vorgesehen. Die Schwellenspannungskennlinie des Transistors wird von der auf dem Schweb-Gate gehaltenen Ladungsmenge gesteuert. Das bedeutet, daß die Mindestspannungsmenge (Schwelle), die an das Steuer-Gate angelegt werden muß, ehe der Transistor "eingeschaltet" wird, damit eine Leitung zwischen seinem Source- und Drainbereich stattfinden kann, vom Ladungspegel am Schweb-Gate  
25 gesteuert wird.

Da das Schweb-Gate einen ganzen Bereich an Ladung halten kann, kann folglich eine EPROM-Speicherzelle auf einen beliebigen Schwellenpegel innerhalb eines Schwellenfensters programmiert werden. Die Größe des Schwellenfensters, welches vom niedrigsten und höchsten  
30 Schwellenpegel des Bausteins begrenzt ist, hängt von den Merkmalen, Betriebsbedingungen und der Vergangenheit des Bausteins ab. Im Prinzip kann jeder deutliche Schwellenpegel innerhalb des Fensters zur Bezeichnung eines bestimmten Speicherzustandes der Zelle herangezogen werden.

Für einen EPROM-Speicher wird der als Speicherzelle dienende Transistor durch Beschleunigen  
35 von Elektronen aus dem Substratkanalbereich durch ein dünnes Gate-Dielektrikum auf das Schweb-Gate auf einen von zwei Zuständen programmiert. Die Speicherzustände sind durch das Entfernen der Ladung auf dem Schweb-Gate mittels ultravioletter Bestrahlung löschar.

Ein EEPROM hat einen ähnlichen Aufbau, bietet aber zusätzlich noch einen Mechanismus zum  
40 Entfernen der Ladung von seinem Schweb-Gate, wenn die richtigen Spannungen angelegt werden. Eine Anordnung derartiger EEPROM-Zellen wird als "Flash-EEPROM array" (Blitz-EEPROM-Anordnung) bezeichnet, wenn eine ganze Anordnung aus Zellen oder eine signifikante Gruppe von Zellen der Anordnung gleichzeitig gelöscht wird (d.h. blitzartig). Eine Zelle kann, sobald sie gelöscht ist, neu programmiert werden.

Eine spezifische einzelne Zelle in einer zweidimensionalen Anordnung aus EPROM, EEPROM-Zellen wird zum Lesen durch das Anlegen einer Source-Drain-Spannung an Source- und Drainleitungen in einer die adressierte Zelle enthaltenden Spalte und Anlegen einer Steuer-Gate-Spannung an eine Wortleitung, die mit den Steuer-Gates in einer die adressierte Zelle enthaltenden Reihe verbunden ist, adressiert.

Der Zustand einer adressierten Speicherzelle eines Transistors wird dadurch gelesen, daß eine Betriebsspannung an seine Source und Drain sowie an sein Steuer-Gate angelegt und dann der Pegel des zwischen der Source und Drain fließenden Stroms erfaßt wird. Der Strompegel ist proportional zum Schwellenpegel des Transistors, der seinerseits von der Ladungsmenge des Schweb-Gates bestimmt ist.

In der üblichen, zwei Zustände aufweisenden EEPROM-Zelle wird ein Trennschwellenpegel festgelegt, um das Schwellenfenster in zwei Bereiche zu unterteilen. Der Source/Drainstrom wird mit dem Trennschwellenpegel verglichen, der beim Programmieren der Zelle benutzt wurde. Liegt der abgelesene Strom höher als der der Schwelle, wird bestimmt, daß sich die Zelle in einem "Null"-Zustand befindet, während bei einem Strom, der die Schwelle nicht erreicht, bestimmt wird, daß sich die Zelle in dem anderen Zustand befindet. Eine derartige Zelle mit zwei Zuständen speichert also ein Bit digitaler Information. Häufig wird als Teil eines Speichersystems zum Erzeugen des Trennschwellenstroms eine Stromquelle benutzt, die von außen programmierbar sein kann.

Für eine Mehrzustands-EEPROM-Speicherzelle speichert also jede Zelle zwei oder mehr Datenbits. Die Information, die eine gegebene EEPROM-Anordnung speichern kann, wird folglich um das Mehrfache der Anzahl von Zuständen erhöht, die jede Zelle speichern kann.

Hauptaufgabe der vorliegenden Erfindung ist es daher, ein System von EEPROM-Speicherzellen zu schaffen, bei dem die Zellen zum Speichern von mehr als einem Bit Daten benutzt werden können.

Aufgabe der Erfindung ist es auch, Leseschaltungen zu schaffen, die einfacher, leichter herzustellen sind und bei verbesserter Genauigkeit und Zuverlässigkeit eine längere Lebensdauer haben.

Aufgabe der vorliegenden Erfindung ist auch die Schaffung verbesserter Programmschaltungen als Teil eines EPROM- oder EEPROM-Speicherchips einer integrierten Schaltung.

Eine Aufgabe der Erfindung besteht auch in der Schaffung von Programmschaltungen, die einfacher, leichter herzustellen und von verbesserter Genauigkeit und Zuverlässigkeit über eine längere Lebensdauer sind.

Noch eine weitere Aufgabe der vorliegenden Erfindung besteht in der Schaffung von Blitz-EEPROM-Halbleiterchips, die Speichervorrichtungen in Form von Magnetplatten in Rechnersystemen ersetzen können.

- 5 Schließlich ist es auch eine Aufgabe der vorliegenden Erfindung, eine Blitz-EEPROM-Konstruktion zu schaffen, die eine erhöhte Lebensdauer, ausgedrückt als Anzahl der Programmier/Lesezyklen, welche der Speicher aushält, haben kann.

#### 10 Zusammenfassung der Erfindung

Diese Ziele werden von der Erfindung wie in den Ansprüchen 1 und 9 bestimmt, erreicht.

- Bei einem Ausführungsbeispiel werden im Lesevorgang unmittelbar die Schwellenpegel in den  
15 Lokalreferenzzellen benutzt, die zuvor von den Hauptreferenzzellen kopiert wurden. Bei einem weiteren Ausführungsbeispiel werden für den Lesevorgang indirekt die Schwellenpegel in den Lokalreferenzzellen benutzt, obwohl das Lesen im Verhältnis zu den Hauptreferenzzellen erfolgt. Das geschieht dadurch, daß zunächst die Lokalreferenzzellen im Verhältnis zu den Hauptreferenzzellen gelesen werden. Die festgestellten Unterschiede werden benutzt, um anschließende,  
20 regelmäßige Lesungen aus Speicherzellen gegenüber den Hauptreferenzzellen zu versetzen, so daß die Lesungen mit Versatz sich effektiv auf die Lokalreferenzzellen beziehen.

Bei einem Ausführungsbeispiel wird der gelesene, durch eine Zelle fließende Strom mit jedem der Schwellenstrompegel der Referenzzellen einer nach dem anderen verglichen.

- 25 Bei einem weiteren Ausführungsbeispiel wird der durch eine Zelle fließende, zu lesende Strom gleichzeitig mit dem des Satzes an Referenzzellen verglichen. Mittels einer speziellen Stromspiegelkonfiguration wird der zu lesende Strom ohne Verschlechterung seines Signals in vielfachen Zweigen reproduziert, einem für jeden Schwellenstromvergleich.

- 30 Gemäß einem weiteren Ausführungsbeispiel der vorliegenden Erfindung werden die Vorgänge des Programmierens und Verifizierens jeweils an einer Menge (d.h. mehreren Bytes) adressierter Zellen durchgeführt. Ferner wird der Verifizierungsvorgang von Schaltungen auf dem EEPROM-Chip durchgeführt. Damit werden Verzögerungen im seriellen Datenversand vom Chip zwecks  
35 Verifizierung zwischen jedem Programmierschritt vermieden.

- Gemäß einem weiteren Ausführungsbeispiel der vorliegenden Erfindung verifiziert eine Schaltung, für den Fall, daß ein programmierter Zustand durch wiederholte Schritte des Programmierens und Verifizierens aus dem "gelöschten" Zustand erhalten wird, den programmierten Zustand nach  
40 jedem Programmierschritt mit dem beabsichtigten Zustand und unterbindet selektiv das weitere Programmieren irgendwelcher der Zellen in der Menge, von denen verifiziert wurde, daß sie korrekt programmiert wurden. Dies ermöglicht ein wirksames paralleles Programmieren einer Menge von Daten bei einer Mehrzustandsausführung.

Gemäß einem weiteren Ausführungsbeispiel der vorliegenden Erfindung wird, für den Fall, daß eine Menge EEPROM-Zellen zur parallelen Löschung adressiert werden, ein gelöschter Zustand durch wiederholte Schritte des Löschsens und Verifizierens aus dem bestehenden Zustand in den "gelöschten" Zustand erhalten, eine Schaltung verifiziert den gelöschten Zustand nach jedem Löschrückschritt mit dem "gelöschten" Zustand und unterbindet selektiv weiteres Löschen irgendwelcher Zellen in der Menge, von denen verifiziert wurde, daß sie korrekt gelöscht wurden. Hiermit wird ein Überlöschen verhindert, welches für den Baustein eine Belastung bedeutet, und wirksames paralleles Löschen einer Gruppe von Zellen wird ermöglicht.

Gemäß einem weiteren Ausführungsbeispiel der vorliegenden Erfindung werden, nachdem eine Gruppe von Zellen durch Löschen in den "gelöschten" Zustand gebracht wurde, die Zellen in den dem "gelöschten" Zustand benachbarten Zustand neu programmiert. Hiermit wird sichergestellt, daß jede gelöschte Zelle aus einem gut bestimmten Zustand beginnt, und außerdem kann jede Zelle eine ähnliche Programmier/Löschbelastung erfahren.

Gemäß einem weiteren Ausführungsbeispiel der vorliegenden Erfindung sind die den Steuer-Gates der EEPROM-Zellen zugeführten Spannungen über einen weiten Bereich hinweg veränderlich und unabhängig von der Spannung, die den Leseschaltungen zugeführt wird. Das erlaubt eine exakte Programmier/Löschrandausnutzung ebenso wie die Verwendung für Prüf- und Diagnosezwecke.

Der hier vorgestellte Gegenstand ist eine Weiterentwicklung der in der Veröffentlichung US 5 095 344 beschriebenen Lesetechniken für EEPROM-Anordnungen.

In US 4 460 982 (Gee et al.) wird ein Programmiersystem offenbart, welches parallel an 8 Bits adressierter Zellen arbeitet. Vor einem Schreibvorgang werden alle 8 Bits der adressierten Zellen gelöscht. Dann wird in die 8 Zellen ein Byte Daten programmiert. Diejenigen unter den 8 adressierten Zellen, die als in den "programmierten" Zustand zu programmieren bezeichnet wurden, werden gleichzeitig Programmierimpulsen unterzogen, bis von allen verifiziert ist, daß sie programmiert wurden.

Von MASUOKA et al. wurde im IEEE JOURNAL OF SOLID-STATE CIRCUITS, Bd. SC-22, Nr. 4, Juni 1987, New York, USA, SS. 548-552 unter "A 256-kbit Flash EEPROM using Triple-Polysilicon Technology" ein Programmier- und Löschrückplan für EEPROM offenbart. Bei dem Löschrückvorgang wird eine Löschrückspannung an alle Löschrück-Gates in der Anordnung angelegt, so daß die ganze Anordnung blitzartig gelöscht werden kann. Ein neuer Leseverstärker nutzt einen Bezugsschwellenpegel, der durch Konkretisieren des Herausziehens der Ausgangsspannung einer unprogrammierten Bezugszelle erzeugt wurde.

Ein EEPROM, der geeignet ist, mehr als zwei Zustände pro Zelle zu speichern, ist offenbart worden im IEEE JOURNAL OF SOLID-STATE CIRCUITS, Band SC-22, Nr. 3, Juni 1987, New York, USA, SS. 460-463 von Bleiker et al.: "A Four-State EEPROM using Floating-Gate Memory Cells". Gleichfalls offenbart sind eine Programmier- und eine Leseschaltung, die an einer Zelle zur Zeit arbeiten. Programmierimpulse werden nacheinander an die Zelle angelegt, und es wird abgeschaltet, wenn die Zelle den gewünschten Zustand erreicht hat.

US 4 691 298 (Fukuda et al.) offenbart einen EEPROM mit einer Latch-Schaltung, die als ein Seriell-Parallel-Umsetzer zum Umwandeln serieller Signale von außen funktioniert. Der Schreibvorgang kann Bit für Bit vorgenommen oder in Hochgeschwindigkeitsweise durchgeführt werden, wenn die umgewandelten, parallelen Signale in den Speicher mit 8 x 8 Bits oder 8 x 4 Bits auf einmal eingegeben werden.

Weitere Ziele, Merkmale und Vorteile der vorliegenden Erfindung werden anhand der nachfolgenden Beschreibung der bevorzugten Ausführungsbeispiele verständlich, wobei die Beschreibung im Zusammenhang mit den beigefügten Zeichnungen zu betrachten ist.

#### Kurzbeschreibung der Zeichnungen

Fig. 1 ist eine Querschnittsansicht einer integrierten Schaltungskonstruktion eines EEPROM-Bausteins, die zum Verwirklichen der verschiedenen Aspekte der vorliegenden Erfindung benutzt werden kann;

Fig. 2 ist eine Ansicht der in Fig. 1 gezeigten Konstruktion im Schnitt 2-2 derselben;

Fig. 3 ist ein Ersatzschaltbild einer einzigen EEPROM-Zelle des in Fig. 1 und 2 veranschaulichten Typs;

Fig. 4 zeigt eine adressierbare Anordnung von EEPROM-Zellen;

Fig. 5 ist ein Blockschaltbild eines EEPROM-Systems, bei dem die verschiedenen Aspekte der vorliegenden Erfindung verwirklicht sind;

Fig. 6 veranschaulicht das Aufteilen des Schwellenfensters einer EEPROM-Zelle, die ein Bit Daten speichert;

Fig. 7A veranschaulicht das Aufteilen des Schwellenfensters einer EEPROM-Zelle, die zwei Bits Daten speichert;

Fig. 7B veranschaulicht das Aufteilen des Schwellenfensters des Source-Drain-Leitungsstroms der in Fig. 7A gezeigten EEPROM-Zelle;

Fig. 8A und 8B sind Kurven, welche die Änderungen und Merkmale eines typischen EEPROM nach einer Nutzungsperiode veranschaulichen;

Fig. 9A veranschaulicht Lese- und Programmierschaltungen für eine Hauptreferenzzelle und eine adressierte Speicherzelle entsprechend einem Ausführungsbeispiel der vorliegenden Erfindung;

Fig. 9B veranschaulicht Mehrzustands-Leseschaltungen mit Referenzzellen gemäß einem Ausführungsbeispiel der vorliegenden Erfindung;

Fig. 9C(1) - 9C(8) veranschaulichen die Zeitsteuerung für Mehrzustandslesen der in Fig. 9B gezeigten Schaltungen;

Fig. 9D veranschaulicht ein Ausführungsbeispiel einer Mehrzustands-Leseschaltung, bei der der Speicherzustand einer Adressenzelle gegenüber einem Satz von Bezugsstrompegeln gleichzeitig erfaßt wird;

Fig. 9E veranschaulicht ein Ausführungsbeispiel einer in Fig. 9D gezeigten IREF-Schaltung als eine EEPROM-Zelle, die mit einem Bezugsstrom programmiert ist;

Fig. 9F veranschaulicht eine bevorzugte Verwirklichung des in Fig. 9D gezeigten Ausführungsbeispiels, bei dem die IREF-Schaltung von einer Stromquelle bereitgestellt wird, die einen in der EEPROM-Zelle programmierten Bezugsstrom reproduziert;

Fig. 9G veranschaulicht ein weiteres Ausführungsbeispiel einer in Fig. 9D gezeigten IREF-Schaltung, bei der ein Bezugsstrom in jedem Zweig der Leitung eines Transistors von vorherbestimmter Größe vorgesehen ist;

Fig. 9H veranschaulicht ein weiteres Ausführungsbeispiel einer Mehrzustands-Leseschaltung, bei der der Speicherzustand einer Adressenzelle gegenüber einem Satz an Bezugsstrompegeln gleichzeitig erfaßt wird;

Fig. 9I veranschaulicht noch ein weiteres Ausführungsbeispiel einer Mehrzustands-Leseschaltung, bei der der Speicherzustand einer Adressenzelle gegenüber einem Satz von Bezugsstrompegeln gleichzeitig erfaßt wird;

Fig. 10 veranschaulicht eine spezifische Speicherorganisation gemäß einem Ausführungsbeispiel der vorliegenden Erfindung;

Fig. 11 zeigt einen Algorithmus zum Programmieren eines Satzes von Lokalreferenzzellen gemäß einem Ausführungsbeispiel der vorliegenden Erfindung;

Fig. 12A zeigt ein Ausführungsbeispiel einer Leseschaltung, die Lokalreferenzzellen direkt benutzt;

Fig. 12B zeigt einen Lesealgorithmus für das Ausführungsbeispiel der Fig. 12A;

Fig. 13A zeigt ein alternatives Ausführungsbeispiel einer Leseschaltung, die Lokalreferenzzellen indirekt benutzt;

Fig. 13B ist eine programmierbare Schaltung für das versetzte Lesen der Hauptreferenzzellen gemäß dem alternativen Ausführungsbeispiel;

Fig. 13C ist ein Detailschaltkreisdigramm für die programmierbare Vorspannschaltung gemäß Fig. 13B;

Fig. 13D zeigt einen Lesealgorithmus für das Ausführungsbeispiel gemäß Fig. 13A;

Fig. 14 veranschaulicht die Lese/Programmierdatenpfade für eine Menge paralleler Zellen;

Fig. 15 zeigt einen Programmier/Verifizieralgorithmus auf einem Chip gemäß der vorliegenden Erfindung;

Fig. 16 ist ein Schaltkreisdigramm für die Vergleichsschaltung der vorliegenden Erfindung;

Fig. 17 ist ein Schaltkreisdigramm für die Programmierschaltung mit Unterbinden gemäß der vorliegenden Erfindung;

Tabelle 1 und 2 sind Listen typischer Beispiele von Betriebsspannungen für die EEPROM-Zelle gemäß der vorliegenden Erfindung.

#### Beschreibung der bevorzugten Ausführungsbeispiele

Es gibt viele spezifische integrierte EPROM-, EEPROM-Halbleiterschaltungsstrukturen, die zur Schaffung einer Speicheranordnung benutzt werden können, mit der die verschiedenen Aspekte der vorliegenden Erfindung in vorteilhafter Weise verwirklicht werden.

#### EEPROM-Zelle mit "geteiltem Kanal"

Ein bevorzugter EEPROM-Aufbau ist insgesamt in den Querschnittsansichten einer integrierten Schaltung gemäß Fig. 1 und 2 dargestellt. Kurz gesagt umfaßt dieser Aufbau zwei Speicherzellen 11 und 13, die auf einem leicht p-dotierten Substrat 15 ausgebildet sind. Ein stark n-dotierter, implantierter Bereich 17 zwischen den Zellen 11 und 13 dient als Drain für die Zelle 1 und als Source für die Zelle 13. Ähnlich ist ein weiterer implantierter, n-dotierter Bereich 19 die Source der Zelle 11 und die Drain einer benachbarten Zelle und ähnlich für einen weiteren n-dotierten Bereich 21.

Jede der Speicherzellen 11 und 13 enthält jeweilige leitfähige Schwebe-Gates 23 und 25, die insgesamt aus Polysiliziummaterial hergestellt sind. Zur gegenseitigen Isolierung und zum Isolieren gegenüber jeglichen weiteren leitfähigen Elementen des Aufbaus ist jedes dieser Schwebe-Gates von dielektrischem Material umgeben. Über die beiden Zellen 11 und 13 erstreckt sich ein Steuer-Gate 27, welches gegenüber den Schwebe-Gates und dem Substrat selbst isoliert ist. Wie Fig. 2 zeigt, sind zusätzlich leitfähige Streifen 29 und 31 isoliert voneinander und gegenüber weiteren



leitfähigen Elementen des Aufbaus vorgesehen und dienen als Lösch-Gates. Ein Paar derartiger Lösch-Gates umgibt das Schwebe-Gate jeder Speicherzelle und ist von ihm durch eine dielektrische Löschschicht getrennt. Die Zellen sind durch Dickschichtoxidbereiche isoliert, zum Beispiel die im Querschnitt in Fig. 1 gezeigten Bereiche 33, 35 und 37 sowie die Bereiche 39 und 41, die in der Ansicht gemäß Fig. 2 gezeigt sind.

Programmiert wird die Speicherzelle durch Übertragen von Elektronen aus dem Substrat 15 an das Schwebe-Gate, beispielsweise das Schwebe-Gate 25 der Speicherzelle 13. Die Ladung am Schwebe-Gate 25 wird durch Elektronen erhöht, die sich aus einem stark p-dotierten Bereich 43 über das Dielektrikum auf das Schwebe-Gate zu bewegen. Entfernt wird Ladung vom Schwebe-Gate durch das Dielektrikum zwischen ihm und den Lösch-Gates 29 und 31. Der bevorzugte EEPROM-Aufbau und ein Verfahren zum Herstellen desselben ist im einzelnen in US 5 070 032 beschrieben.

Bei dem in Fig. 1 und 2 gezeigten EEPROM-Aufbau handelt es sich um einen Typ mit "geteiltem Kanal". Jede Zelle kann als zusammengesetzter Transistor betrachtet werden, der aus zwei in Reihe liegenden Transistoren T1 und T2 besteht, wie Fig. 3 zeigt. Der Transistor T1 ist ein entlang der Länge L1 des Kanals der Zelle 11 gemäß Fig. 1 ausgebildeter Transistor 11a. Er hat eine veränderliche Schwellenspannung  $V_{T1}$ . In Reihe mit dem T1 Transistor 11a liegt der T2 Transistor 11b, der in einem Teil des Kanals L2 ausgebildet ist. Dieser hat eine feste Schwellenspannung  $V_{T2}$  von etwa 1 V. Elemente der Ersatzschaltung gemäß Fig. 3 sind mit den gleichen Bezugszeichen versehen wie für die entsprechenden Teile in Fig. 1 und 2 benutzt, wobei ein Apostroph hinzugefügt ist.

Wie am besten aus der Ersatzschaltung gemäß Fig. 3 zu entnehmen ist, beeinflusst das Ladungsniveau am Schwebe-Gate 23' des T1 einer EEPROM-Zelle die Schwellenspannung  $V_{T1}$  des T1 Transistors 11a bei einer Betätigung mit dem Steuer-Gate 27'. So läßt sich in einer Zelle eine Anzahl von Speicherzuständen bilden, die gut bestimmten Schwellenspannungen entsprechen, welche durch geeignete, dem Schwebe-Gate auferlegte Ladungsmengen in die Zelle programmiert werden. Das Programmieren geschieht durch Anlegen zweckmäßiger Spannungen über eine gewisse Zeit hinweg an das Steuer-Gate 27' ebenso wie die Source 17' und die Drain 19' der Zelle.

#### Adressierbare Blitz-EEPROM-Anordnung

Die verschiedenen Aspekte der vorliegenden Erfindung werden typischerweise bei einer Anordnung aus Blitz-EEPROM-Zellen in einem integrierten Schaltungsmikrobaustein angewandt. Eine Anordnung einzeln adressierbarer EEPROM-Zellen 60 ist in Fig. 4 schematisch dargestellt. Jede Zelle gleicht der in Fig. 3 gezeigten einen Zelle und hat ein Steuer-Gate, Source und Drain sowie ein Lösch-Gate. Die Vielzahl einzelner Speicherzellen ist in Reihen und Spalten organisiert. Jede Zelle wird durch wahlweises gleichzeitiges Erregen ihrer Reihe und Spalte adressiert. Eine Spalte 62 umfaßt zum Beispiel eine erste Speicherzelle 63, eine benachbarte zweite Speicherzelle 65 und so weiter. Eine zweite Spalte 72 umfaßt Speicherzellen 73, 75 und so weiter. Die Zellen 63

und 73 liegen in einer Reihe 76, die Zellen 65 und 71 in einer weiteren, benachbarten Reihe und so weiter.

Längs jeder Reihe ist eine Wortleitung an alle Steuer-Gates der Zellen in der Reihe angeschlossen. Zum Beispiel hat die Reihe 76 die Wortleitung 77 und die nächste Reihe die Wortleitung 79. Ein Reihendekodierer 81 schließt wahlweise die Steuer-Gatespannung  $V_{CG}$  auf einer Eingabeleitung 83 an alle Steuer-Gates längs einer ausgewählten Wortleitung für eine Reihe an.

Längs einer Spalte sind die Sourcen aller Zellen mittels einer Sourceleitung, beispielsweise 91 und die Drains aller Zellen mittels einer Drainleitung, beispielsweise 93 verbunden. Da die Zellen längs einer Spalte über ihre Sourcen und Drains in Reihe geschaltet sind, ist die Drain einer Zelle auch die Source der benachbarten Zelle. Die Leitung 93 ist also die Drainleitung für die Spalte 62 und auch die Sourceleitung für die Spalte 72. Ein Spaltendekodierer 101 schließt wahlweise die Sourcespannung  $V_S$  auf einer Eingabeleitung 103 an alle Sourcen und die Drainspannung  $V_D$  auf einer Eingabeleitung 105 an alle Drains längs einer ausgewählten Spalte an.

Jede Zelle wird von der Reihe und Spalte adressiert, in der sie sich befindet. Wenn zum Beispiel die Zelle 75 zum Programmieren oder Lesen adressiert wird, müssen dem Steuer-Gate, der Source und Drain der Zelle geeignete Programmier- oder Lesespannungen zugeführt werden. Eine Adresse auf dem internen Adressenbus 111 wird zum Dekodieren des Reihendekodierers 81 benutzt, um  $V_{CG}$  an die Wortleitung 79 anzuschließen, die mit dem Steuer-Gate der Zelle 75 verbunden ist. Die gleiche Adresse wird zum Dekodieren des Spaltendekodierers 101 benutzt, um  $V_S$  an die Sourceleitung 93 und  $V_D$  an die Drainleitung 95 anzuschließen, die mit der Source bzw. Drain der Zelle 75 verbunden sind.

Ein Aspekt der vorliegenden Erfindung, der in einem Absatz weiter unten mehr im einzelnen offenbart wird, ist die Verwirklichung des parallelen Programmierens und Lesens einer Vielzahl von Speicherzellen. Um eine Vielzahl von Spalten gleichzeitig auszuwählen, steuert der Spaltendekodierer seinerseits das Schalten eines Sourcemultiplexers 107 und eines Drainmultiplexers 109. Auf diese Weise können die Sourceleitungen und Drainleitungen der ausgewählten Vielzahl von Spalten zum Anschluß an  $V_S$  bzw.  $V_D$  zugänglich gemacht werden.

Der Zugang zum Lösch-Gate jeder Zelle ist ähnlich wie der zum Steuer-Gate. Bei einer Verwirklichung wird eine Löschleitung, beispielsweise 113 oder 115 oder 117 mit dem Lösch-Gate jeder der Zellen in einer Reihe verbunden. Ein Löschdekodierer 119 dekodiert eine Adresse auf dem internen Adressenbus 111 und schließt wahlweise die Löschspannung  $V_{EG}$  auf der Eingabeleitung 121 an eine Löschleitung an. Das macht es möglich, jede Reihe der Zellen unabhängig zu adressieren, beispielsweise die Reihe 76, die gleichzeitig (blitzartig) durch geeignete Spannungen gelöscht wird, welche an ihre Lösch-Gates über die Löschleitung 113 angelegt werden. In diesem Fall besteht die Blitzzelle aus einer Reihe von Speicherzellen. Es sind aber auch andere Verwirklichungen von Blitzzellen möglich, und in den meisten Anwendungsfällen wird für gleichzeitiges Löschen vieler Reihen von Zellen auf einmal gesorgt.

### Blitz-EEPROM-System

Die in Fig. 4 gezeigte adressierbare EEPROM-Anordnung 60 bildet einen Teil des größeren Mehrzustands-Blitz-EEPROM-Systems der vorliegenden Erfindung, wie es in Fig. 5 dargestellt ist.

In dem größeren System wird ein EEPROM-IC-Chip 130 von einem Controller 140 über eine Schnittstelle 150 gesteuert. Der Controller 140 selbst steht in Verbindung mit einer zentralen Mikroprozessoreinheit 160.

Der EEPROM-Chip 130 weist die adressierbare EEPROM-Anordnung 60, eine serielle Protokollverknüpfung 170, lokale Leistungssteuerschaltungen 180 sowie verschiedene Programmier- und Leseschaltungen 190, 200, 210, 220, 230 und 240 auf.

Der Controller 140 steuert das Wirken des EEPROM-Chips 130, indem er die geeigneten Spannungen, Steuerungen und Synchronisierung liefert. In den Tabellen 1 und 2 sind typische Beispiele von Spannungsbedingungen für die verschiedenen Betriebsweisen der EEPROM-Zelle gezeigt. Die adressierbare EEPROM-Anordnung 60 kann von dem Controller 140 direkt betrieben oder, wie in Fig. 5 gezeigt, auf dem Chip weiter mittels der lokalen Leistungssteuerung 180 geregelt werden. Steuer- und Datenverbindungen zwischen dem Controller 140 und dem Chip 130 erfolgen über die Leitung seriell EIN 251 und die Leitung seriell AUS 253. Über Leitung 255 bietet der Controller eine Taktung.

In einem typischen Betriebsfall des EEPROM-Chips 130 sendet der Controller 140 einen seriellen Signalstrom über die Leitung seriell EIN 251 an den Chip 130. Die Steuer-, Daten-, Adressen- und Zeitinformation enthaltenden Signale werden von der seriellen Protokollverknüpfung 170 sortiert. In angemessener Zeitfolge gibt die Verknüpfung 170 verschiedene Steuersignale 257 zum Steuern der verschiedenen Schaltungen auf dem Chip 130 aus. Sie sendet außerdem eine Adresse über den internen Adressenbus 111, um die adressierte Zelle an vom Controller ausgegebene Spannungen anzuschließen. Wenn es sich bei der Betätigung um Programmieren handelt, werden in der Zwischenzeit Daten zum Programmieren der adressierten Zelle aufbereitet, indem sie über eine serielle Datenleitung 259 an einen Satz von Lese/Programmier-Latch-Gliedern und Schieberegister 190 geschickt werden.

### Leseschaltungen und Techniken unter Verwendung von Referenzzellen

Den Speicherzustand einer Zelle exakt und zuverlässig festzustellen ist für den EEPROM-Einsatz wesentlich. Das liegt daran, daß alle grundlegenden Funktionen, wie Lesen, Verifizieren des Löschsens und Verifizieren der Programmierung davon abhängen. Verbesserte und neue Leseschaltungen 220 für den EEPROM-Chip 130 und Techniken gemäß der vorliegenden Erfindung machen Mehrzustands-EEPROM möglich.

Wie schon im Zusammenhang mit Fig. 3 erwähnt, bestimmt die auf das Schwebegate 23' gelegte, programmierte Ladung die programmierte Schwellenspannung  $V_{T1}$  der Zelle. Insgesamt nimmt  $V_{T1}$  zu oder ab mit der Menge der negativen Ladung am Schwebegate 23'. Die Ladung kann sogar auf einen positiven Wert verringert werden (Verarmungsmodus), wenn  $V_{T1}$  bis unter

$V_{T1}$  absinkt und sogar negativ wird. Die maximalen und minimalen Werte für  $V_{T1}$  hängen von der dielektrischen Stärke des Materials des Bausteins ab. Die Spanne von  $V_{T1}$  definiert ein Schwellenspannungsfenster, in welchem Speicherzustände verwirklicht werden können.

US 5 095 344 offenbart eine EEPROM-Zelle mit Speicherzuständen, die innerhalb eines maximierten Fensters der Schwellenspannung  $V_{T1}$  bestimmt sind. Das volle Schwellenspannungsfenster umschließt den negativen Bereich der Schwellenspannung zusätzlich zu dem üblichen positiven Bereich. Das vergrößerte Fenster bietet mehr Speicherraum zum Verwirklichen des Mehrzustands in einer EEPROM-Zelle.

Die Figuren 6 und 7 zeigen die Art und Weise, in der das Schwellenspannungsfenster für eine Speicherzelle mit einem Zweizustandsspeicher bzw. einem Vierzustandsspeicher aufgeteilt ist. (Natürlich ist es auch möglich, das Fenster für einen Dreizustandsspeicher oder sogar für ein Kontinuum an Zuständen in einem analogen statt digitalen Speicher aufzuteilen.)

Unter Hinweis zunächst auf Fig. 6 zeigt die durchgezogene Kurve 343  $V_{T1}$  als eine Funktion der Programmierzeit. Das Schwellenspannungsfenster ist von dem Minimal- und Maximalwert von  $V_{T1}$  begrenzt, die etwa von dem Löschzustandspegel 345 bzw. dem Vollprogrammierzustandspegel 347 wiedergegeben sind. Der Zweizustandsspeicher ist durch die Aufteilung des Fensters in zwei Hälften 346, 348 mit Hilfe eines Trennschwellenpegels 349 verwirklicht. So läßt sich sagen, daß die Zelle sich im Speicherzustand 0 (oder Zustand 1) befindet, wenn die Zelle mit  $V_{T1}$  innerhalb des Bereichs 346 (bzw. innerhalb des Bereichs 348) programmiert ist.

Ein typischer Löschen/Programmierzklus beginnt mit dem Löschen, durch das die Schwellenspannung der Zelle auf ihren Löschzustandspegel 345 herabgesetzt wird. Anschließendes wiederholtes Programmieren wird angewandt, um die Schwellenspannung  $V_{T1}$  auf das gewünschte Niveau anzuheben. Statt während einer gewissen, festgelegten Zeitspanne entsprechend dem Zustand, auf den die Zelle programmiert werden soll, Programmierspannungen kontinuierlich an die adressierte Zelle anzulegen, wird bevorzugt, Programmierspannungen in wiederholten, kurzen Impulsen anzulegen und nach jedem Impuls einen Lesevorgang erfolgen zu lassen, um festzustellen, wann die Programmierung auf den gewünschten Schwellenspannungspegel erreicht ist. In diesem Zeitpunkt endet das Programmieren. Die Programmierspannungen und die Dauer der Impulse sind so gewählt, daß die Impulse  $V_{T1}$  schnell über die verschiedenen Bereiche vorwärts-treiben, jeder Impuls dabei aber so fein ist, daß er über keinen der Bereiche hinauschießt. Hierdurch werden spannungs- und feldabhängige Beanspruchungen der Zelle auf ein Minimum eingeschränkt und folglich ihre Zuverlässigkeit verbessert.

In Fig. 7A ist der Vierzustandsfall dargestellt, bei dem das Schwellenspannungsfenster durch Trennpegel 352, 354 bzw. 356 in vier Bereiche 351, 353, 355, 357 aufgeteilt ist. Die Zelle wird als im Zustand "3" oder "2" oder "1" oder "0" befindlich betrachtet, wenn ihre Spannung  $V_{T1}$  so programmiert ist, daß sie innerhalb der entsprechenden Bereiche 351 bzw. 353 bzw. 355 bzw. 357 liegt. Eine Vierzustandszelle kann zwei Datenbits speichern. Die vier Zustände können also als (1,1), (1,0), (0,1) bzw. (0,0) kodiert werden.

Allgemein gesagt muß das Schwellenfenster, wenn jede EEPROM-Zelle  $K$  Zustände speichern soll, in  $K$  Bereiche mit mindestens  $K-1$  Schwellenpegeln aufgeteilt werden. Es ist also nur ein Trennpegel für eine Zweizustandsspeicherzelle nötig, während für eine Vierzustandszelle drei Trennpegel erforderlich sind.

Im Prinzip kann ein Schwellenspannungsfenster in eine große Anzahl von Speicherzuständen aufgeteilt sein. Für einen EEPROM-Baustein mit einem maximalen Schwellenfenster von 16 V kann es zum Beispiel in 32 Zustände aufgeteilt sein, jedes innerhalb eines Intervalls von etwa einem halben Volt. In der Praxis speichern bisher bekannte EEPROM-Bausteine nur zwei Zustände oder ein Bit pro Zelle mit verringerter Zuverlässigkeit und Lebensdauer. Abgesehen davon, daß sie mit einem kleineren Schwellenfenster arbeiten, bewältigen bekannte Bausteine auch nicht zwei weitere Probleme, die EEPROM-Bausteinen innewohnen. Beide Probleme beziehen sich auf die Ungewißheit hinsichtlich der Ladungsmenge im Schwebegate und infolgedessen die Ungewißheit über die in die Zelle programmierte Schwellenspannung  $V_{T1}$ .

Das erste Problem hat mit der die Dauerhaftigkeit beeinflussenden Beanspruchung zu tun, die der Baustein immer dann erfährt, wenn er einen Löscho/Programmierzyklus durchläuft. Die Dauerhaftigkeit eines Blitz-EEPROM-Bausteins ist seine Fähigkeit, einer gegebenen Anzahl von Programmier/Löschozyklen standzuhalten. Das physikalische Phänomen, welches die Dauerhaftigkeit bekannter Blitz-EEPROM-Bausteine begrenzt, ist das Einfangen von Elektronen in den aktiven dielektrischen Schichten der Vorrichtung. Während des Programmierens werden Elektronen aus dem Substrat durch eine dielektrische Zwischenfläche in das Schwebegate injiziert. Ähnlich werden während des Löschoens Elektronen aus dem Schwebegate durch eine dielektrische Zwischenschicht in das Löscho-Gate extrahiert. In beiden Fällen bleiben einige der Elektronen in der dielektrischen Zwischenschicht gefangen. Diese eingefangenen Elektronen stehen dem angelegten elektrischen Feld in anschließenden Programmier/Löschozyklen entgegen und verursachen dadurch, daß die programmierte  $V_{T1}$  auf einen niedrigeren Wert und die gelöschte  $V_{T1}$  auf einen höheren Wert verlagert wird. Das zeigt sich in einem allmählichen Schließen im Spannungsfenster zwischen dem "0" und dem "1" Zustand bekannter Vorrichtungen, wie in Fig. 8A gezeigt. Jenseits von etwa  $1 \times 10^4$  Programmier/Löschozyklen kann das Verschließen des Fensters solche Ausmaße annehmen, daß die Leseschaltungsanordnung nicht mehr richtig funktioniert. Werden dann die Zyklen fortgesetzt, erleidet die Vorrichtung letztendlich einen katastrophalen Ausfall wegen eines Bruchs des Dielektrikums. Das geschieht typischerweise bei zwischen  $1 \times 10^6$  und  $1 \times 10^7$  Zyklen und ist als struktureller Zusammenbruch der Vorrichtung bekannt. Bei bekannten EEPROM-Bausteinen ist das Schließen des Fensters der Grund für die Begrenzung der praktischen Dauerhaftigkeit auf etwa  $1 \times 10^4$  Programmier/Löschozyklen. Dieses Problem erlangt eine noch kritischere Bedeutung, wenn ein Mehrzustandsspeicher verwirklicht wird, denn dafür ist eine genauere Platzierung von  $V_{T1}$  erforderlich.

Eine zweite Schwierigkeit hat mit dem Beibehalten der Ladung am Schwebegate zu tun. Es besteht die Tendenz, daß die Ladung am Schwebegate durch Leckverluste im Verlauf der Zeit etwas abnimmt. Das hat zur Folge, daß die Schwellenspannung  $V_{T1}$  sich auch im Lauf der Zeit auf einen niedrigeren Wert verlagert. Fig. 8B veranschaulicht die Abnahme von  $V_{T1}$  als Funktion der Zeit. Während der Lebensdauer der Vorrichtung kann  $V_{T1}$  sich um bis zu 1 V verlagern. In

einer Mehrzustandsvorrichtung könnte dies den Speicher um einen oder zwei Zustände verschieben.

Mit der vorliegenden Erfindung werden diese Probleme überwunden, und es werden Schaltkreise und Techniken geboten, mit denen die verschiedenen Zustände zuverlässig programmiert und gelesen werden können, und zwar sogar bei einer Mehrzustandsverwirklichung.

Der Speicherzustand einer Zelle kann durch Messen der darin programmierten Schwellenspannung  $V_{T1}$  bestimmt werden. Als Alternative kann der Speicherzustand, wie in US 5 095 344 beschrieben, zweckmäßigerweise dadurch bestimmt werden, daß die unterschiedliche Leitung im Source-Drainstrom  $I_{DS}$  für die verschiedenen Zustände gemessen wird. Für das Beispiel des Zustandes "1" zeigt Fig. 7A die Aufteilung im Schwellenspannungsfenster. Fig. 7B hingegen zeigt typische Werte für  $I_{DS}$  (durchgezogene Kurven) für die vier Zustände als Funktion der Steuer-Gate-Spannung  $V_{CG}$ . Wenn  $V_{CG}$  bei 5 V liegt, können die  $I_{DS}$  Werte für jeden der vier Leitungszustände dadurch unterschieden werden, daß mit vier entsprechenden Stromleseverstärkern parallel abgetastet wird. Jedem Verstärker ist ein entsprechender Bezugspegel  $I_{REF}$  für die Leitungszustände zugeordnet (als gestrichelte Kurven in Fig. 7B gezeigt). So wie die Trennschwellenpegel (siehe Fig. 6 und 7A) zum Abgrenzen der unterschiedlichen Bereiche im Schwellenspannungsfenster benutzt werden, werden die  $I_{REF}$  Pegel benutzt, um das gleiche in dem entsprechenden Fenster des Source-Drainstroms zu tun. Der Leitungszustand der Speicherzelle kann bestimmt werden durch einen Vergleich mit den Werten für  $I_{REF}$ . In US 5 095 344 wird vorgeschlagen, sowohl für das Programmieren als auch das Lesen die gleichen Leseverstärker und  $I_{REF}$  Werte zu benutzen. Damit ist eine gute Nachführung zwischen den Bezugspegeln (gestrichelte Kurven in Fig. 89) und den programmierten Pegeln (durchgezogene Kurven in Fig. 7B) möglich.

Bei dem verbesserten System der vorliegenden Erfindung werden die  $I_{REF}$  Werte selbst von den Source-Drainströmen eines Satzes von EEPROM-Zellen geliefert, die auf dem gleichen Chip existieren und allein für diesen Zweck reserviert sind. Sie wirken also als Hauptreferenzzellen, und ihre  $I_{REF}$  Werte werden als Bezugspegel für das Lesen und Programmieren aller weiteren EEPROM-Zellen auf dem gleichen Chip herangezogen. Wenn man die gleiche Vorrichtung heranzieht wie die EEPROM-Zellen, um als Referenzzellen zu wirken, wird eine ausgezeichnete Nachführung hinsichtlich Schwankungen der Temperatur, Spannung und Verarbeitung erzielt. Ferner wird das Problem des Beibehaltens der Ladung, was für die Mehrzustandsverwirklichung wichtig ist, gemildert.

In Fig. 9A ist eine solche Hauptreferenzzelle 400 mit ihren Programmier- und Lesepfaden gezeigt. Das Referenzzellenlösch- und Programmiermodul 410 dient zum Programmieren und Neuprogrammieren jeder dieser Referenzzellen 400. Das Modul 410 umfaßt Programmier- und Löschschaltungen 411 mit einem Programmierpfad 413, der an den Drain der Hauptreferenzzelle 400 angeschlossen ist. Die Schaltungen 411 werden von Adressen initiiert, die aus dem internen Bus 111 mittels eines Programmdekodierers 415 bzw. eines Löschdekodierers 417 dekodiert werden. Dementsprechend werden Programmierspannungen oder Löschspannungen wahlweise an jede Referenzzelle geliefert, beispielsweise Zelle 400. Auf diese Weise kann der Bezugspegel in jeder

Referenzzelle unabhängig gesetzt oder umprogrammiert werden. Üblicherweise ist der Schwellenpegel jeder Referenzzelle vom Hersteller auf den für jede Charge hergestellter Chips angemessenen, optimalen Pegel programmiert. Das könnte durch einen Vergleich mit einem externen, genormten Bezugspegel geschehen. Der Benutzer hat mittels Softwaresteuerung die Wahl, die  
 5 Bezugschwellenpegel neu zu setzen.

Sobald die Bezugsschwellenspannung  $V_{T1}$  oder der Bezugssource-Drainstrom  $I_{REF}$  in jede Referenzzelle 400 programmiert worden ist, dient dies zur Bezugnahme für das Lesen einer adressierten Speicherzelle, beispielsweise der Zelle 420. Die Referenzzelle 400 ist mit einem  
 10 ersten Schenkel 403 eines Stromleseverstärkers 410 über einen getakteten Schalter 413 verbunden. Ein zweiter Schenkel 415 des Verstärkers ist im wesentlichen mit der adressierten Speicherzelle 420 verbunden, deren programmierter Leitungszustand festgestellt werden soll. Wenn die Zelle 420 gelesen werden soll, wird ein Schalter 421 mittels eines Steuersignals LESEN angesteuert, so daß der Drain der Zelle mit dem zweiten Schenkel 415 verbunden wird. Der  
 15 Leseverstärker 410 liefert Spannung über  $V_{CC}$  an die Drains sowohl der Hauptreferenzzelle 400 als auch der adressierten Zelle 420. Bei dem bevorzugten Ausführungsbeispiel hat der Verstärker eine Stromspiegelkonfiguration, so daß jegliche Differenz in den durch die beiden Schenkel 403 und 415 fließenden Strömen dazu führt, daß die Spannung im zweiten Schenkel 415 hinauf zu  $V_{CC}$  oder hinab zu  $V_S$  gezogen wird. Folglich ist der Knotenpunkt am zweiten Schenkel 415  
 20 jeweils HOCH (oder TIEF), wenn der Source-Drainstrom  $I_{DS}$  für die adressierte Zelle 420 weniger (oder mehr) ist als  $I_{REF}$  durch die Hauptreferenzzelle 400. In einem geeigneten, von einem getakteten Schalter 423 gesteuerten Zeitpunkt kann das erfaßte Ergebnis am zweiten Schenkel 415 mittels eines Latch-Gliedes 425 gehalten und an einer Ausgabeleitung 427 zur Verfügung gestellt werden. Ist  $I_{DS}$  weniger als  $I_{REF}$ , erscheint an der Ausgabeleitung 427 HOCH, und die  
 25 adressierte Zelle 420 wird als im gleichen Leitungszustand wie die Hauptreferenzzelle 400 betrachtet.

Bei dem bevorzugten Ausführungsbeispiel ist zwischen dem zweiten Schenkel 415 und dem Drain 431 der adressierten Zelle 420 außerdem eine Schaltung 430 zum Spannungsblockieren und Schnellhochziehen eingesetzt. Diese Schaltung 430 dient dazu, die Drainspannung  $V_D$  auf  
 30 einem Maximum von 1,5 V bis 2,0 V zu halten, wenn im Fall eines niedrigeren  $I_{DS}$  hochgeladen wird. Außerdem verhindert sie, daß  $V_D$  zu tief herabgezogen wird, wenn  $I_{DS}$  höher ist.

Wenn jede Speicherzelle K Zustände speichern soll, müssen, allgemein gesagt, mindestens K-1  
 35 oder vorzugsweise K Bezugspegel vorgesehen werden. Bei einem Ausführungsbeispiel wird die adressierte Zelle mit den K Referenzzellen unter Verwendung von parallelen k Leseverstärkern verglichen. Dies wird für den Fall mit zwei Zuständen wegen der Geschwindigkeit bevorzugt, kann aber den verfügbaren Strom zu sehr "strecken", als daß im Mehrzustandsfall eine ordentliche Wahrnehmung möglich wäre. Für den Mehrzustandsfall wird also vorzugsweise die adressierte Zelle mit den K Referenzzellen jeweils einzeln der Reihe nach verglichen.  
 40

Fig. 9B zeigt die Mehrzustandslesekonfiguration ausdrücklicher. Die K Referenzzellen, beispielsweise 431, 433, 435 sind mit dem Leseverstärker 440 über den ersten Schenkel 441 des Verstärkers verbunden. Die Verbindung wird von getakteten Schaltern, beispielsweise 451, 453

bzw. 455 zeitmultiplexiert. Der zweite Schenkel 457 des Leseverstärkers ist mit der adressierten Zelle verbunden, wie in Fig. 9A. Das erfaßte Signal am zweiten Schenkel 457 wird von getakten Schaltern, beispielsweise 461, 463, 465 zeitselektiv in Latch-Glieder 471, 473, 475 eingerastet.

5

In den Fig. 9C(1)-9C(8) ist die Taktung für das Mehrzustandslesen veranschaulicht. Wenn das Signal LESEN auf HOCH geht, wird ein Schalter 421 angesteuert und die adressierte Speicherzelle mit dem zweiten Schenkel 457 des Leseverstärkers 440 verbunden (Fig. 9C(1)). Die Zeitbestimmung der Takte ist in den Fig. 9C(2)-9C(4) angegeben. Bei jedem Taktsignal vergleicht  
 10 also der Leseverstärker die adressierte Zelle der Reihe nach mit jeder der Referenzzellen und rastet jedes Ergebnisse ein. Die eingerasteten Ergebnisse des Leseverstärkers sind in den Fig. 9C(5)-9C(7) angegeben. Wenn alle K Ausgangszustände des Leseverstärkers 440 eingerastet sind, werden sie von einem K-L Dekodierer 480 ( $2^L > K$ ) zu L binären Bits kodiert (Fig. 9C(8)).

15

Die Mehrschwellenpegel werden also von einem Satz Speicherzellen bereitgestellt, der als Hauptreferenzzellen dient. Die Hauptreferenzzellen sind unabhängig und von außen löschar und programmierbar, und zwar entweder vom Hersteller oder Benutzer des Bausteins. Dieses Merkmal bietet die größtmögliche Flexibilität und erlaubt es, Trennschwellen innerhalb des Schwellenfensters des Bausteins jederzeit individuell zu setzen. Weil es sich um die gleiche Vorrichtung wie die  
 20 der Speicherzellen handelt, ist den Referenzzellen eine dichte Verfolgung der gleichen Schwankungen aufgrund von Herstellungsverfahren, Betriebsbedingungen und Problemen mit dem Beibehalten der Spannung möglich. Die unabhängige Programmierbarkeit jedes Schwellenpegels nach Wunsch erlaubt eine Optimierung und Feineinstellung der Aufteilung des Schwellenfensters, was den Mehrzustandsspeicher realisierbar macht. Ferner ist, je nach Bedarf des Benutzers oder  
 25 den Eigenschaften des Bausteins, im betreffenden Zeitpunkt eine Auslegung auf einen Zweizustands- oder Mehrzustandsspeicher aus dem gleichen Baustein nach der Fertigung möglich.

Ein weiteres Ausführungsbeispiel der vorliegenden Erfindung bietet eine verbesserte Mehrzustandserfassung einer adressierten Speicherzelle. Wie im Zusammenhang mit einem früheren Ausführungsbeispiel zum Erfassen eines Mehrzustandsspeichers schon beschrieben, wird  
 30 vorzugsweise der Leitungsstrom der Zelle gleichzeitig oder parallel mit sämtlichen Referenzpegeln des Leitungsstroms (Schwellenpegeln) verglichen. Eine Vierzustandsspeicherzelle hat zum Beispiel mindestens drei Referenzstrompegel zum Abgrenzen der vier Zustände. Ein paralleles Erfassen des Zustands der Zelle bedeutet einen gleichzeitigen Vergleich des Leitungsstroms der Zelle I<sub>CELL</sub> gegenüber jedem der drei Referenzstrompegel. Das ist schneller als ihn der Reihe nach mit jedem  
 35 der drei Referenzleitungspegel zu vergleichen. Bei dem zuvor beschriebenen einfacheren Ausführungsbeispiel könnte allerdings der Leitungsstrom der adressierten Zelle dadurch verdünnt werden, daß er in drei Zweige, einen für jeden Referenzpegelvergleich unterteilt wird. Eine einfache Verwirklichung der gleichzeitigen oder parallelen Mehrzustandserfassung kann sich also  
 40 wegen des Rauschabstandserfordernisses des Erfassungssystems verbieten, insbesondere wenn es sich um viele Zustände handelt.

Die Fig. 9D - 9I zeigen verschiedene Ausführungsbeispiele der gleichzeitigen Mehrzustandserfassung ohne den Nachteil einer Verschlechterung des Leitungsstroms der erfaßten Zelle. In jedem



Ausführungsbeispiel wird eine Stromspiegelschaltung Eins-zu-Viele verwendet, um einen Strom in vielen Kopien zu reproduzieren, so daß jede Kopie zum Vergleichen mit einem Referenzstrompegel zur gleichen Zeit benutzt werden kann.

In Fig. 9D ist ein erstes Ausführungsbeispiel gleichzeitiger Mehrzustandserfassung dargestellt. Eine Eins-zu-Viele Stromspiegelschaltung weist einen ersten Transistor 910 an einem ersten Schenkel 920 und einen zweiten Transistor 911, 912 ... bzw. 915 an jedem Zweig 921, 922 ... 925 eines zweiten Schenkels auf. Sobald ein erster Strom im ersten Schenkel 920 fließt, verhält sich der zweite Transistor an jedem Zweig des zweiten Schenkels wie eine Stromquelle und liefert in seinem Zweig einen reproduzierten Strom. Das Verhältnis zwischen reproduziertem Strom und dem ersten Strom entspricht einer Skala in Übereinstimmung mit den relativen Größen des zweiten Transistors 911, 912 ... 915 gegenüber dem ersten Transistor 910.

Im vorliegenden Ausführungsbeispiel haben alle Transistoren die gleiche Größe, wie mit dem Symbol "X" in Fig. 9D angemerkt. Das führt zu einer Stromspiegelschaltung Eins-zu-Viele, in der der erste Strom im ersten Schenkel 920 in allen Zweigen 921, 922 ... 925 des zweiten Schenkels identisch reproduziert wird. Wenn also der Leitungsstrom  $I_{CELL}$  einer adressierten Speicherzelle 420 durch einen Lesefreigabeschalter 421 im ersten Schenkel 920 fließt, wird der gleiche Strom  $I_{CELL}$  in den Zweigen 921, 922 ... 925 des zweiten Schenkels reproduziert. Das wird ohne Verdünnung von  $I_{CELL}$  erreicht.

Sobald  $I_{CELL}$  in jedem Zweig reproduziert wurde, wird ein Vergleich mit einem zugeordneten Referenzstrompegel vorgenommen. Dies geschieht, indem jeder Zweig gleichfalls mit einer mit der ersten Stromquelle 911, 912 ... 915 in Reihe liegenden zweiten Stromquelle 931, 932 ... bzw. 935 angesteuert wird. Jede zweite Stromquelle oder  $I_{REF}$ -Schaltung 931, 932 ... 935 liefert jeweils den vorherbestimmten Bezugsstrompegel, beispielsweise  $I_{REF1}$  in Leitung 941 des ersten Zweiges,  $I_{REF2}$  in Leitung 942 des zweiten Zweiges ... und  $I_{REFk}$  in Leitung 953 des kten Zweiges. Der Speicherzustand wird dann durch Erfassen der Lokalisierung des  $I_{CELL}$ -Pegels gegenüber den  $I_{REF}$ -Werten bestimmt. Die für jeden Zustand erfaßten Ausgaben, in Fig. 9D mit SA1, SA2 ... SAK bezeichnet, werden jeweils von einem Knoten 951 des ersten Zweiges, einem Knoten 952 des zweiten Zweiges ... und einem Knoten 953 des kten Zweiges abgeleitet. Der Knoten in jedem Zweig befindet sich zwischen der ersten und zweiten Stromquelle. Insgesamt haben die beiden Stromquellen entgegengesetzte Polarität. Wenn die zweite Stromquelle 931, 932 ... 935 ein an einem Ende mit  $V_S$  verbundener n-Kanal Transistor ist, dann ist die erste Stromquelle ein p-Kanaltransistor 911, 912 ... 915, der am anderen Ende an  $V_{CC}$  angeschlossen ist. Der Knoten wird je nach den relativen Pegeln von  $I_{CELL}$  und  $I_{REF}$  in den beiden Stromquellen entweder nach oben zu  $V_{CC}$  hochgezogen (typischerweise 5 V) oder herab zu  $V_S$  (typischerweise 0 V). Im ersten Zweig wird zum Beispiel ein Strom  $I_{CELL}$  in einer Leitung 921 reproduziert, und ein Strom  $I_{REF1}$  wird in einer Leitung 941 geliefert. Der Knoten 951 ist jeweils HOCH (oder TIEF), wenn  $I_{CELL}$  größer ist als (oder kleiner als)  $I_{REF1}$ . Somit wäre bei einem Speicherzustand, bei dem  $I_{CELL}$  zwischen  $I_{REF1}$  und  $I_{REF2}$  liegt, nur der Knoten 951 HOCH, womit das Resultat eine Mehrzustandsausgabe wäre (SA1, SA2 ... SAK) = (0, 1 ... 1).

Insgesamt kann jede  $I_{REF}$ -Schaltung 931, 932 ... 935 eine Stromquellschaltung sein, die im voraus so eingestellt ist, daß sie die verschiedenen Referenzstrompegel  $I_{REF1}$ ,  $I_{REF2}$  ...  $I_{REF3}$  bietet.

In Fig. 9E ist ein Ausführungsbeispiel bei EEPROM-Anwendungen gezeigt, bei dem jede  $I_{REF}$ -Schaltung 931, 932 ... 935 jeweils von einer Referenzzelle 431, 432 ... 435 bereitgestellt ist, die ihrerseits selbst eine EEPROM-Zelle ähnlich der im Zusammenhang mit Fig. 9A und 9B beschriebenen ist. Das macht die Referenzzelle anwendbar als Hauptreferenzzelle oder Lokalreferenzzelle, in der ein Referenzleistungsstrompegel programmiert werden kann.

In Fig. 9F ist eine bevorzugte Verwirklichung dargestellt, bei der jede  $I_{REF}$ -Schaltung nicht direkt von einer Referenzzelle bereitgestellt wird, sondern statt dessen von einer Reproduktion derselben. Das macht es möglich, eine Menge (zum Beispiel 64) Speicherzellen sich die gleiche Referenzzelle zum gleichzeitigen Erfassen teilen zu lassen. Ein Transistor 961, 962 ... bzw. 965 in jeder der  $I_{REF}$ -Schaltungen 931, 932 ... 935 dient als Stromquelle für die Lieferung des reproduzierten Referenzstrompegels aus jeder der Referenzzellen 431, 432 ... 435. Jeder Transistor wird an seinem Gate von einer Referenzspannung  $REF1$ ,  $REF2$  ...  $REFk$  gesteuert, um die erforderlichen Referenzstrompegel  $I_{REF1}$ ,  $I_{REF2}$  ...  $I_{REF3}$  zu erzeugen. Jede Referenzspannung wird von einer REF-Schaltung 971 ... 975 geliefert. Eine alternative Ansicht sieht vor, daß jeder Transistor 961, 962 ... 965 und die zugeordnete REF-Schaltung 971 ... 975 eine doppelte Stromspiegelschaltung bilden, von der der Referenzstrom jeder Referenzzelle 431, 432 ... 435 als Leitungsstrom des Transistors 961, 962 ... 965 reproduziert wird. Wenn man die  $I_{REF1}$  Schaltung 931 als repräsentativ betrachtet, weist sie den Transistor 961 als eine Stromquelle für  $I_{REF1}$  auf. Der  $I_{REF1}$ -Pegel wird als eine Reproduktion des Leitungsstroms der Referenzzelle 431 erhalten. Die Referenzzelle 431 liefert einen Referenzstrom  $I_{REF1}$  an einen ersten Schenkel 976 der ersten Stromspiegelschaltung, welcher in einem zweiten Schenkel 977 derselben reproduziert wird. Der zweite Schenkel 977 der ersten Stromspiegelschaltung ist mit einem ersten Schenkel der zweiten Stromspiegelschaltung gegenseitig verbunden. Damit wird der reproduzierte Referenzstrom seinerseits im zweiten Schenkel 941 des zweiten Spiegels mittels des Transistors 961 reproduziert. Insgesamt haben die beiden Stromspiegelschaltungen entgegengesetzte Polarität. Wenn zum Beispiel die REF1-Zelle 431 ein n-Kanaltransistor ist, weist die erste Stromspiegelschaltung zwei p-Kanaltransistoren 981 und 982 von gleicher Größe "X" auf, und die zweite Stromspiegelschaltung weist zwei n-Kanaltransistoren 983 und 961 von gleicher Größe "W" auf.

In Fig. 9G ist ein weiteres Ausführungsbeispiel dargestellt, bei dem die verschiedenen, von der zweiten Stromquelle jedes Zweiges gelieferten  $I_{REF}$ -Pegel alle von einer Referenzschaltung 976 erzeugt werden. Die Referenzschaltung 976 stellt eine Referenzspannung zur Verfügung, die an jedes Gate des Transistors 961, 962 ... bzw. 965 jedes Zweiges angelegt wird. Wie bei dem in Fig. 9F gezeigten Ausführungsbeispiel dient die Referenzspannung dazu, die Transistoren einzuschalten. Die unterschiedlichen  $I_{REF}$ -Pegel über die Zweige hinweg werden nunmehr aber durch Einstellen der Größe der Transistoren 961, 962 ... 965 erhalten. Wie zum Beispiel in Fig. 9G gezeigt, haben die Transistoren 961, 962 bzw. 965 Größen  $I*W$ ,  $J*W$  ...  $K*W$ , wo  $I:J: \dots :K$  jeweils in den gleichen Verhältnissen zueinander stehen wie  $I_{REF1}:I_{REF2}: \dots :I_{REFk}$ . Die einzige

Referenzschaltung 976 kann eine Quelle konstanter Spannung oder eine Schaltung mit einer Referenzzelle ähnlich der REF-Schaltung 971 in Fig. 9F sein. Dies gilt unter der normalen Stromspiegelschaltungsbedingung, gemäß der die Transistoren in jedem Zweig, zum Beispiel M81 und 961 in Sättigungsrichtung vorgespannt sind.

In Fig. 9H ist ein weiteres Ausführungsbeispiel dargestellt, bei dem alle zweiten Stromquellen über die Zweige hinweg die gleichen sind, aber  $I_{CELL}$  von der ersten Stromquelle in jeden Zweig hinein mit Pegeln reproduziert wird, die entsprechend der Abstufung der Referenzstrompegel skaliert sind. Das Skalieren erfolgt durch Anpassen der Größe jedes zweiten Transistors 911, 912 ... 915. Wie Fig. 9H zeigt, haben beispielsweise die zweiten Transistoren 911, 912 ... bzw. 915 Größen von  $I \cdot X$ ,  $J \cdot X$  ...  $K \cdot X$ , worin  $X$  die Größe des ersten Transistors 910 im ersten Schenkel 920 ist und  $I:J: \dots :K$  jeweils in den gleichen Verhältnissen zueinander stehen wie  $I_{REF1}:I_{REF2}: \dots :I_{REFK}$ . Es wird also nur eine REF-Schaltung 976 über die Zweige hinweg benutzt, und außerdem sind die Größen aller Transistoren 961, 962 ... 965 nunmehr identisch. Die einzige Referenzschaltung 976 kann eine Quelle konstanter Spannung oder eine Schaltung sein, die eine Referenzzelle ähnlich der in Fig. 9F gezeigten REF-Schaltung 971 beinhaltet. Bei einer Verwirklichung ist die Referenzschaltung 976 so gestaltet, daß jede zweite Stromquelle 961, 962 ... 965 einen Strom liefert, der dem höchsten Referenzstrompegel  $I_{REFK}$  gleicht. Im Verhältnis zu den in Fig. 9D - 9G gezeigten Ausführungsbeispielen ist die Reihenfolge der Ausgaben der Knoten umgekehrt.

Noch ein weiteres Ausführungsbeispiel für gleichzeitige Mehrzustandserfassung mit einer Schaltung ähnlich der in Fig. 9G gezeigten ist in Fig. 9I dargestellt, außer daß die Identitäten der Adressenspeicherzelle und der  $I_{REF}$ -Schaltung ausgewechselt sind. Anders ausgedrückt, in jedem Zweig liefert die zweite Stromquelle, beispielsweise 931, 932 ... 935 nunmehr einen reproduzierten  $I_{CELL}$ . Das wird erreicht mittels einer adressierten Speicherzellenschaltung 977, welche jedem Gate des Transistors 961, 962 ... bzw. 965 jedes Zweigs eine Referenzspannung  $MC$  zuführt. Die Schaltung 977 ähnelt der REF1-Schaltung 971 in Fig. 9F, außer daß die REF1-Zelle 431 nunmehr durch die adressierte Speicherzelle 420 ersetzt ist. In ähnlicher Weise liefert die erste Stromquelle, wie 911, 912 ... 915 nun jeweils  $I_{REF1}$ ,  $I_{REF2}$  ...  $I_{REFK}$ . Die verschiedenen  $I_{REF}$ -Werte werden durch eine skalierte Reproduktion des Stroms einer IREF0-Schaltung 978 erhalten. Das Skalieren erfolgt mittels Einstellen der Größe jedes zweiten Transistors 911, 912 ... 915 in der Stromspiegelschaltung Eins-zu-Viele. Wie in Fig. 9I gezeigt, haben beispielsweise die zweiten Transistoren 911, 912, ... bzw. 915 Größen von  $I \cdot X$ ,  $J \cdot X$  ...  $K \cdot X$ , worin  $X$  die Größe des ersten Transistors 910 im ersten Schenkel 920 ist und  $I:J: \dots :K$  jeweils in den gleichen Verhältnissen zueinander stehen wie  $I_{REF0}:I_{REF1}:I_{REF2}: \dots :I_{REFK}$ . Insgesamt kann die IREF0-Schaltung 978 eine beliebige Stromquelle sein, die einen Strompegel  $I_{REF0}$  liefert. Bei einem Ausführungsbeispiel ist die IREF0-Schaltung eine EEPROM-Zelle, die mit einem Referenzstrompegel programmierbar ist, ähnlich wie im Zusammenhang mit den Fig. 9A und 9B beschrieben.

Ein weiteres wichtiges Merkmal der vorliegenden Erfindung dient dem Überwinden der Schwierigkeiten der mit der Dauerwechselfestigkeit in Beziehung stehenden Beanspruchung. Wie schon erwähnt, hängen die Lösch-, Programmier- und Leseigenschaften jeder Speicherzelle von der kumulierten Beanspruchung ab, welche die Zelle im Verlauf einer Anzahl von Programmier/

Löschzyklen zu durchlaufen hat. Insgesamt werden die Speicherzellen viel mehr Programmier/Löschzyklen unterzogen als die Hauptreferenzzellen. Die Anfangs optimierten Referenzpegel verlieren schließlich ihre Ausrichtung und verursachen Lesefehler. Der gegenwärtige, grundlegende erfinderische Gedanke besteht darin, die Referenzpegel die gleiche zyklische Erfahrung wiedergeben zu lassen, wie sie die Speicherzellen erleiden. Das wird erreicht durch die Verwirklichung von Lokalreferenzzellen zusätzlich zu den Hauptreferenzzellen. Die Lokalreferenzzellen werden den gleichen Programmier/Löschzyklen ausgesetzt wie die Speicherzellen. Jedes Mal nach einem Löschvorgang werden die Referenzpegel in den Hauptreferenzzellen neu in den entsprechenden Satz Lokalreferenzzellen kopiert. Speicherzellen werden dann in Bezug auf die Referenzpegel der dicht nachlaufenden Lokalreferenzzellen gelesen. Auf diese Weise wird die Abweichung in Zellenmerkmalen nach jedem Programmier/Löschzyklus automatisch ausgeglichen. Deshalb wird die richtige Aufteilung des Transformierschwellenfensters beibehalten, so daß die Speicherzustände selbst nach vielen Zyklen noch korrekt gelesen werden können.

Fig. 10 zeigt die Verwirklichung mit örtlichen Zellen zur Bezugnahme für Blitz-EEPROM. In der Blitz-EEPROM-Anordnung 60 (Fig. 4) wird jede Speicherzellengruppe, die kollektiv gelöscht oder programmiert wird, als Sektor bezeichnet. Der Ausdruck "Blitz-Sektor" ist analog dem für Speichervorrichtung in Form von Magnetplatten benutzten Ausdruck "Sektor", und sie werden hier austauschbar verwendet. Die EEPROM-Anordnung ist in Blitz-Sektoren, wie 501, 503 und 505 gruppiert. Während alle Speicherzellen in einem Blitz-Sektor die gleichen Zyklen durchlaufen, können in verschiedenen Blitz-Sektoren verschiedene Zyklen angewandt werden. Um jeden Blitz-Sektor richtig zu verfolgen, wird ein Satz Speicherzellen in jedem Blitz-Sektor zur Verwendung als Lokalreferenzzellen reserviert. Nachdem zum Beispiel der Blitz-Sektor 503 gelöscht wurde, werden die Referenzpegel der Hauptreferenzzellen 507 in die Lokalreferenzzellen umprogrammiert, die dem Blitz-Sektor 503 zugeordnet sind. Bis zum nächsten Löschzyklus fahren die Leseschaltungen 513 fort, die Speicherzellen innerhalb des Blitz-Sektors 503 in Bezug auf die umprogrammierten Referenzpegel zu lesen.

Fig. 11(1)-11(7) veranschaulichen den Algorithmus zum Neuprogrammieren der Referenzzellen eines Sektors. Insbesondere beziehen sich die Fig. 11(1)-11(3) auf das Löschen der Lokalreferenzzellen des Sektors auf ihre "gelöschten Zustände". In Fig. 11(1) wird also ein Löschspannungsimpuls an alle Speicherzellen des Sektors, einschließlich der Lokalreferenzzellen angelegt. In Fig. 11(2) werden dann alle Lokalreferenzzellen in Bezug auf die Hauptreferenzzellen gelesen, um zu verifizieren, ob sie alle auf den "gelöschten Zustand" gelöscht worden sind. Solange noch eine Zelle in einem anderen Zustand vorgefunden wird, wird an sämtliche Zellen ein weiterer Impuls Löschspannung angelegt. Dieses Verfahren wird so lange wiederholt, bis von allen Lokalreferenzzellen im Sektor verifiziert wird, daß sie sich in dem "gelöschten" Zustand befinden (Fig. 11(3)).

Die Fig. 11(4)-11(7) beziehen sich auf das Programmieren der Lokalreferenzzellen im Sektor. Nachdem von allen Lokalreferenzzellen im Sektor verifiziert wurde, daß sie sich im "gelöschten" Zustand befinden, wird ein Impuls Programmiervoltage in Fig. 11(4) nur an alle Lokalreferenzzellen angelegt. Hierauf folgt in Fig. 11(5) das Lesen der Lokalreferenzzellen in Bezug auf die Hauptreferenzzellen, um zu verifizieren, ob jede der Lokalreferenzzellen auf den gleichen Zustand wie die entsprechende Hauptreferenzzelle programmiert ist. Für diejenigen Lokalreferenzzellen,

von denen das nicht verifiziert wird, wird ein weiterer Impuls Programmierspannung selektiv an diese allein angelegt (Fig. 11(6)). Dieses Verfahren wird so lange wiederholt, bis von allen Lokalreferenzzellen korrekt verifiziert wurde (Fig. 11(7)), daß sie auf die verschiedenen Trennschwellenpegel im Schwellenfenster programmiert sind.

Sobald die Lokalreferenzzellen im Sektor neuprogrammiert wurden, werden sie direkt oder indirekt zur Löscheckverifizierung, Programmverifizierung oder zum Lesen der adressierten Speicherzellen des Sektors benutzt.

Fig. 12A zeigt ein Ausführungsbeispiel, bei dem die Lokalreferenzzellen direkt zum Lesen oder Programmier/Löscheckverifizieren der Speicherzellen des Sektors benutzt werden. Während dieser Vorgänge wird ein paralleles Paar Schalter 525 von einem LESE-Signal angesteuert, und der Leseverstärker 440 liest die adressierten Speicherzellen 523 des Sektors in Bezug auf jede der Lokalreferenzzellen 525 des Sektors. Während der Programmier/Löscheckverifizierung der Lokalreferenzzellen (wie in Fig. 11 dargestellt), gibt ein weiteres paralleles Paar Schalter 527 das Lesen der Lokalreferenzzellen 525 in Bezug auf die Hauptreferenzzellen 529 frei.

Fig. 12B veranschaulicht den Algorithmus für den direkten Gebrauch der Lokalreferenzzellen zum Lesen oder Programmier/ Löscheckverifizieren der adressierten Speicherzellen des Sektors.

Fig. 13A zeigt ein alternatives Ausführungsbeispiel, bei dem die Lokalreferenzzellen indirekt zum Lesen der adressierten Speicherzelle benutzt werden. Zunächst werden die Hauptreferenzzellen gelöscht und je auf eine der gewünschten mehrfachen Trennschwellen innerhalb des Schwellenfensters programmiert. Unter Verwendung dieser Hauptreferenzschwellen werden die Lokalreferenzzellen innerhalb eines gelöschten Sektors von Zellen je auf eine der gleichen, gewünschten, mehrfachen Trennschwellen programmiert. Als nächstes werden die adressierten Zellen im Sektor programmiert mit den gewünschten Daten (beschrieben). An der Lesefolge für die adressierten Zellen im Sektor sind dann die in Fig. 13A gezeigten Schritte beteiligt.

Zuerst wird jede der Lokalreferenzzellen 525 in Bezug auf die entsprechende Hauptreferenzzelle 531 gelesen. Das geschieht durch ein freigebendes Signal LESEN 1 an einen Schalter 533, der die Lokalreferenzzellen 525 mit dem zweiten Schenkel 457 des Leseverstärkers 440 mit der Hauptreferenz 531 verbindet, die mit dem ersten Schenkel 441 des Leseverstärkers verbunden ist. Hilfsstromquellschaltungen, die jeder Hauptreferenzzelle zugeordnet sind, werden nunmehr benutzt, um den Strom durch den ersten Schenkel 441 des Leseverstärkers optimal vorzuspannen, damit er zu dem Strom im zweiten Schenkel 457 paßt. Wenn der Vorspanneinstellvorgang für alle Trennschwellenpegel beendet ist, werden die adressierten Zellen im Sektor in Bezug auf die vorspannungsmäßig eingestellten Hauptreferenzzellen gelesen. Dies geschieht durch Sperren von LESEN 1 an 533 und Freigeben des Signals LESEN an den Schalter 535. Der Vorteil dieses Ansatzes besteht darin, daß jegliche Schwankungen des Wertes  $V_{CC}$ , der Temperatur, zyklische Ermüdung oder sonstige Wirkungen, die im Verlauf der Zeit Schwellenabweichungen zwischen den Hauptreferenzzellen und den adressierten Zellen hervorrufen können, vor dem Lesen eliminiert wird, da die Lokalreferenzzellen (die Schwellenabweichungen der adressierten Zellen verfolgen), benutzt werden, um die Trennschwellen der Hauptreferenzzellen wirksam neu

einzustellen. Dieses System erlaubt zum Beispiel das Programmieren der adressierten Zellen, wenn die Hauptreferenzzellen mit  $V_{CC}=5,5$  V Leistung versorgt werden und anschließend das Lesen der adressierten Zellen, wenn die Hauptreferenzzellen bei  $V_{CC}=4,5$  V betrieben werden. Die Differenz von 1 Volt in  $V_{CC}$ , die normalerweise zu einer Änderung des Wertes der Trennschwellen führen würde, wird dadurch neutralisiert, daß die Lokalreferenzzellen zur Vorspannungseinstellung der Hauptreferenzzellen herangezogen werden, um dieser Änderung im Zeitpunkt des Lesens entgegenzuwirken.

Die Fig. 13B und 13C zeigen mehr im einzelnen ein Ausführungsbeispiel der Stromvorspannungsschaltungen, wie 541, 543, 545 für die Hauptreferenzzellen 551, 553, 555. Jede Vorspannungsschaltung wirkt als Stromnebenschluß für den Strom in der Hauptreferenzzelle. Die Schaltung 541 zum Beispiel greift vom Drain der Hauptreferenzzelle 551 über die Leitung 561 ab. Sie modifiziert den Strom in der Leitung 562 zum Leseverstärker (erster Schenkel), entweder durch Sourcестrom von  $V_{CC}$  oder Drainstrom zu  $V_{SS}$ . Im ersten Fall wird der Strom in der Leitung 562 verringert und im zuletzt genannten Fall umgekehrt. Da das Vorspannen für die Hauptreferenz 551 festgelegt wird, kann jegliche Ungleichheit in den Strömen in den beiden Schenkeln des Leseverstärkers vom Chip nach außen mitgeteilt werden. Dies wird vom Controller (siehe Fig. 5) festgestellt, der seinerseits die Vorspannungsschaltung 541 über den internen Adressenbus 111 programmiert, um Strom in der Leitung 562 zu subtrahieren oder zu addieren, um den der lokalen Referenz auszugleichen.

Fig. 13C zeigt ein Ausführungsbeispiel der Vorspannungsschaltung, beispielsweise der Schaltung 541. Eine Reihe paralleler Transistoren, wie 571, 573, 575 sind alle mit ihren Drains an  $V_{CC}$  angeschlossen und mit ihren Sources über Schalter, beispielsweise 581, 583, 585 an die Leitung 561. Durch selektives Ansteuern der Schalter können Transistoren in unterschiedlicher Zahl benutzt werden, um von der Leitung 562 verschiedene Menge Strom zu subtrahieren. Ähnlich ist eine weitere Reihe paralleler Transistoren, zum Beispiel 591, 593, 595 alle mit ihren Sources an  $V_{SS}$  angeschlossen und mit ihren Drains über Schalter, wie 601, 603, 605 an die Leitung 561. Durch selektives Ansteuern der Schalter kann eine verschiedene Anzahl Transistoren benutzt werden, um verschiedene Menge Strom zur Leitung 562 zu addieren. Ein Dekodierer 609 wird benutzt, zum Adressendekodieren vom internen Adressenbus 111, um die Schalter selektiv anzusteuern. Die freigebenden Signale sind in Latch-Gliedern 611, 613 gespeichert. Auf diese Weise werden die Hauptreferenzzellen, jedes Mal wenn ein Sektor gelesen wird, gegenüber den Lokalreferenzzellen neu vorgespannt und zum Lesen der Speicherzellen im Sektor benutzt.

Die Fig. 13D(1)-13D(4) veranschaulichen den Lesealgorithmus für das alternative Ausführungsbeispiel. Die Lokalreferenzzellen des Sektors müssen zuvor in Bezug auf die Hauptreferenzzellen programmiert und verifiziert worden sein (Fig. 13D(1)). Dementsprechend wird dann jede Hauptreferenzzelle in Bezug auf die Lokalreferenzzellen gelesen (Fig. 13D(2)). Die Hauptreferenzzellen werden vorgespannt, um den Strom auszugleichen auf den der entsprechenden Lokalreferenzzellen (Fig. 13D(3)). Anschließend werden die Speicherzellen im Sektor in Bezug auf die vorgespannten Hauptreferenzzellen gelesen (Fig. 13D(4)).

Die Leseschaltungen und der beschriebene Betrieb wird gleichfalls beim Programmieren und Löschen der Speicherzellen angewandt, insbesondere im verifizierenden Teil des Vorganges. Wie zuvor beschrieben, erfolgt die Programmierung in kleinen Schritten, wobei der programmierte Zustand zwischendurch gelesen wird, um zu verifizieren, ob der gewünschte Zustand erreicht wurde. Sobald der programmierte Zustand als korrekt verifiziert wurde, hört das Programmieren auf. Ähnlich wird das Löschen in kleinen Schritten durchgeführt, wobei der Löschzustand zwischendurch gelesen wird, um zu verifizieren, ob der "gelöschte" Zustand erreicht worden ist. Sobald der "gelöschte" Zustand richtig verifiziert wird, hört das Löschen auf.

Wie schon beschrieben, sind nur K-1 Trennschwellenpegel erforderlich, um das Schwellenfenster in K Bereiche aufzuteilen, was der Speicherzelle das Speichern von K Zuständen erlaubt. Gemäß einem Aspekt der vorliegenden Erfindung ist es jedoch im Fall des Mehrzustands, bei dem das Schwellenfenster feiner aufgeteilt ist, vorzuziehen, K Schwellenpegel für K Zustand zu benutzen. Der zusätzliche Schwellenpegel wird benutzt, um den "gelöschten" Zustand von demjenigen Zustand zu unterscheiden, der den niedrigsten Schwellenpegel hat. Das verhindert ein Überlöschen und damit ein Überbeanspruchen der Zelle, da das Löschen aufhört, sobald der "gelöschte" Zustand erreicht ist. Die selektive Hemmung individueller Zellen für das Löschen gilt nicht für den Blitz-EEPROM-Fall, wo mindestens ein Sektor jedes Mal gelöscht werden muß. Geeignet ist sie für jene EEPROM-Anordnungen, wo die Speicherzellen zum Löschen individuell adressiert werden können.

Gemäß einem weiteren Merkmal der Erfindung wird eine Speicherzelle, nachdem sie auf den "gelöschten" Zustand gelöscht wurde, geringfügig programmiert, um die Zelle in den Zustand mit dem niedrigsten Schwellenpegel (Grundzustand) dem "gelöschten" Zustand benachbart) zu bringen. Das hat zwei Vorteile. Zunächst sind die Schwellenpegel des Grundzustands aller Speicherzellen, da sie zwischen den beiden gleichen Trennschwellenpegeln eingegrenzt sind, gut bestimmt und nicht breit gestreut. Damit ergibt sich ein einheitlicher Ausgangspunkt für das anschließende Programmieren der Zellen. Zweitens erhalten alle Zellen eine gewisse Programmierung, was verhindert, daß jene Zellen, bei denen die Tendenz besteht, daß in ihnen der Grundzustand gespeichert ist, zum Beispiel den Anschluß an die übrigen verlieren, was den Durchlauf der Programmier/Löschzyklen und den Hintergrund für die Dauerfestigkeit betrifft.

#### Programmverifizierung auf dem Chip

Wie zuvor erwähnt, wird das Programmieren einer EEPROM-Zelle in einen gewünschten Zustand vorzugsweise in kleinen Schritten, ausgehend von dem "Löschen"-Zustand vorgenommen. Nach jedem Programmierschritt wird die in der Programmierung befindliche Zelle gelesen, um zu verifizieren, ob der gewünschte Zustand erreicht worden ist. Ist das nicht der Fall, wird weiteres Programmieren und Verifizieren wiederholt, bis die Verifizierung das ergibt.

Unter Hinweis auf das in Fig. 5 dargestellte Systemdiagramm steht der EEPROM-Chip 130 unter der Steuerung durch den Controller 140. Sie sind durch die Leitung seriell EIN 251 und die Leitung seriell AUS 253 miteinander seriell verbunden. Bei bekannten EEPROM-Bausteinen wird nach jedem Programmierschritt der Zustand, der in der im Programmierprozeß befindlichen Zelle

erreicht wurde, gelesen und an den Controller 140 oder die CPU 160 zwecks Verifizierung mit dem gewünschten Zustand zurückgeschickt. Dieses System bringt einen großen Nachteil mit sich für die Geschwindigkeit, insbesondere in Anbetracht der seriellen Verbindung.

Bei der vorliegenden Erfindung ist die Programmverifizierung dadurch optimiert, daß eine Menge (typischerweise mehrere Bytes) Zellen parallel programmiert wird, worauf ein paralleles Verifizieren auf dem Chip erfolgt. Das parallele Programmieren wird durch eine selektive Programmierschaltung verwirklicht, die das Programmieren jener Zellen in der Menge sperrt, deren Zustände bereits richtig verifiziert wurden. Hierbei handelt es sich um ein wesentliches Merkmal bei einer Mehrzustandsverwirklichung, denn einige Zellen erreichen ihren gewünschten Zustand früher als andere und fahren fort über den gewünschten Zustand, wenn sie nicht angehalten werden. Nachdem die ganze Menge Zellen richtig verifiziert wurde, wird dieser Zustand durch Verknüpfung auf dem Chip an den Controller mitgeteilt, wodurch die Programmierung der nächsten Menge Zellen beginnen kann. Auf diese Weise braucht Information nicht zwischen jedem Programmierschritt zwischen dem EEPROM-Chip und dem Controller hin- und hergeschickt zu werden, und die Programmierverifizierungsgeschwindigkeit wird stark gefördert.

Fig. 14 veranschaulicht die Programmier- und Verifizierpfade für eine Menge  $n$  paralleler Zellen. Die gleichen Bezugsziffern werden für entsprechende Module im Systemdiagramm der Fig. 5 benutzt. Die EEPROM-Anordnung 60 wird von  $N$  Zellen auf einmal adressiert.  $N$  kann zum Beispiel 64 Zellen breit sein. In einem 512 Byte Blitzsektor, der aus 4 Reihen zu 1024 Zellen besteht, gibt es 64 Mengen zu 64 Zellen. Der Sourcemultiplexer 107 schließt wahlweise die  $N$  Quellen einer adressierten Menge Zellen an die Sourcespannung  $V_S$  auf Leitung 103 an. Auf ähnliche Weise macht der Drainmultiplexer 109 wahlweise die  $N$  Drains der Menge durch einen  $N$ -Kanaldatenpfad 105 zugänglich. Auf den Datenpfad 105 wird von der Programmierschaltung beim Programmieren mit Inhibieren 210 Zugriff genommen und beim Lesen, Programmverifizieren oder Löschverifizieren mit Leseschaltungen 220.

Um noch einmal auf das Systemdiagramm in Fig. 5 zurückzukommen, die Programmierung steht unter der Steuerung durch den Controller 140. Die in den Sektor zu programmierende Information wird Menge für Menge gesandt. Der Controller sendet zuerst eine erste Menge  $N \cdot L$  serieller Datenbits zusammen mit Adressen, Steuer- und Zeitinformation an den EEPROM-Chip 130.  $L$  ist die Anzahl binärer Bits, die pro Speicherzelle kodiert sind. Zum Beispiel  $L = 1$  für eine 2-Zustandszelle und  $L = 2$  für eine 4-Zustandszelle. Wenn also  $N = 64$  und  $L = 2$ , ist die Menge der Datenbits 128 Bits breit. Die  $N \cdot L$  Datenbits werden in den Latch-Gliedern und Schieberegistern 190 gespeichert, wo die seriellen Bits in  $N \cdot L$  parallele Bits umgewandelt werden. Diese Daten sind erforderlich für die Programmverifizierung im Zusammenhang mit den Leseschaltungen 220, dem Bitdekodierer 230, der Vergleichsschaltung 200 und der Programmierschaltung mit Inhibieren 210.

Der Programmalgorithmus für eine Menge  $N$  Zellen läßt sich am besten unter Hinweis sowohl auf das Systemdiagramm der Fig. 5 und die Figuren 15(1)-15(7) beschreiben, die den Algorithmus selbst veranschaulichen. Wie in einem früheren Abschnitt erwähnt, muß der gesamte Sektor, ehe er programmiert wird, gelöscht werden, und von allen Zellen in ihm verifiziert werden, daß sie



sich in dem "gelöschten" Zustand befinden (Fig. 15(1)). Hierauf folgt in Fig. 15(2) das Programmieren der Lokalreferenzzellen des Sektors (wie in Fig. 11(1)-(3) gezeigt. In Fig. 15(3) werden die  $N \cdot L$  Bits paralleler Daten in Latch-Glieder 190 eingerastet. In Fig. 15(4) nehmen die Leseschaltungen 220 Zugriff auf den N-Kanaldatenpfad 105, um die Zustände in der N Menge Zellen zu lesen. Der Lesealgorithmus ist bereits im Zusammenhang mit Fig. 12B oder 13D beschrieben worden. Die N Zell-Lesungen erzeugt  $N \cdot K$  ( $K =$  Anzahl Zustände pro Zelle) Ausgabezustände. Diese werden vom Bitdekodierer 230 zu  $N \cdot L$  binären Bits dekodiert. In Fig. 15(5) werden die  $N \cdot L$  gelesenen Bits von der Vergleichsschaltung 200 Bit für Bit mit den  $N \cdot L$  Programmierdatenbits aus Latch-Gliedern 190 verglichen. Gemäß Fig. 15(6) wird, wenn irgend ein gelesenes Bit sich nicht mit dem Programmierdatenbit vergleichen läßt, ein weiterer Programmierspannungsimpuls von der Programmierschaltung 210 gleichzeitig an die Menge Zellen angelegt. Allerdings blockiert eine Inhibierschaltung innerhalb der Programmierschaltung 210 selektiv das Programmieren an jene Zellen, deren Bits mit den programmierten Datenbits richtig verifiziert sind. Folglich werden nur die nicht verifizierten Zellen jedes Mal programmiert. Das Programmieren und Verifizieren wird wiederholt, bis alle Zellen korrekt verifiziert sind gemäß Fig. 15(7).

Fig. 16 zeigt ein Ausführungsbeispiel der Vergleichsschaltung 200 aus Fig. 5 mehr im einzelnen. Die Schaltung 200 weist N Zellenvergleichsmodule, wie 701, 703 auf, eines für jede der N Zellen in der Menge. In jedem Zellenvergleichsmodul, wie dem Modul 701, werden die L gelesenen Bits ( $L =$  Anzahl der für jede Zelle kodierten binären Bits) Bit für Bit mit den entsprechenden Programmierdatenbits verglichen. Dies wird von L XOR-Gattern, wie 711, 713, 715 vorgenommen. Die Ausgabe dieser XOR-Gatter durchlaufen ein NOR-Gatter 717, so daß am Ausgang des NOR-Gatters immer dann eine "1" erscheint, wenn alle L Bits verifiziert sind, und eine "0" erscheint, wenn es anders ist. Wenn das Steuersignal VERIFIZIEREN wahr ist, wird dieses Ergebnis in ein Latch-Glied 721 eingerastet, so daß das gleiche Ergebnis am Ausgang des NOR-Gatters 717 am Ausgang 725 des Zellenvergleichsmoduls bereitsteht. Die Vergleichsschaltung 200 führt die Vergleiche von L Bits parallel durch. Die Ausgaben der N Vergleichsmodule, wie 725, 727 sind an einer N-Kanalausgabeleitung 731 verfügbar, um der Programmierschaltung mit Inhibieren 210 gemäß Fig. 5 zugeführt zu werden.

Zur gleichen Zeit werden die N Ausgaben, wie 725, 727 durch ein UND-Gatter 733 geleitet, so daß dessen einzige Ausgabe 735 als eine "1" resultiert, wenn alle N Zellen verifiziert sind und als eine "0", wenn es anders ist. Unter Hinweis auch auf Fig. 5 wird die einzige Ausgabe 735 dazu benutzt, dem Controller 140 zu signalisieren, daß alle N Zellen in der Menge Daten richtig verifiziert wurden. Das Signal in der Ausgabe 735 wird während eines Verifizierungsvorganges durch die Leitung seriell AUS 253 über UND-Gatter 240 geschickt.

Beim Hochfahren oder am Ende des Programmier/Verifizierens einer Menge Daten werden alle Ausgaben der Zellenvergleichsmodule, wie 725, 727 auf den "nichtverifizierten" Zustand von "0" zurückgesetzt. Das wird erreicht, indem der Knoten 726 mittels des RÜCKSETZ-Signals in Leitung 727 an einen Transistor 729 auf  $V_{SS}$  gezogen wird (0 V).

Fig. 17 zeigt ein Ausführungsbeispiel der Programmierschaltung mit Inhibieren 210 aus Fig. 5 mehr im einzelnen. Die Programmierschaltung 210 weist N Programmier-mit-Inhibier-Module, wie

801, 803 auf. Wie in Tabelle 1 und 2 dargestellt, muß zum Programmieren der N Zellen eine Spannung  $V_{PD}$  an jede der Drains der N Zellen und eine Spannung  $V_{PG}$  an die Steuer-Gates angelegt werden. Jedes Programmiermodul, wie 801, dient zum selektiven Weiterleiten von  $V_{PD}$  auf einer Leitung 805 über den einen der N Kanaldatenpfad 105 an eine der Drains. Da  $V_{PD}$  üblicherweise etwa 8 V bis 9 V ist, was höher ist als  $V_{CC}$ , kann die letztere Spannung nicht benutzt werden, um den Transistorschalter 807 einzuschalten. Stattdessen wird die höhere Spannung  $V_{CG}$  (etwa 12 V) zum Ansteuern des Schalters 807 benutzt.  $V_{CG}$  in der Leitung 801 selbst wird von einem UND-Gatter freigegeben, wenn sowohl das Programmiersteuersignal PGM auf einer Leitung 813 wahr und das Signal auf einer Leitung 731 eine "0" ist. Da das Signal auf der Leitung 731 vom Ausgang des in Fig. 16 gezeigten Zellenvergleichsmoduls 701 kommt, folgt daraus, daß  $V_{PD}$  selektiv zu jenen Zellen weitergeleitet wird, die noch nicht verifiziert sind. Auf diese Weise wird jedes Mal, wenn ein Programmierimpuls angelegt wird, dieser nur an jene Zellen angelegt, die ihre beabsichtigten Zustände noch nicht erreicht haben. Dieses Merkmal selektiver Programmierung ist besonders nötig bei der Verwirklichung des parallelen Programmierens und der Verifizierung auf dem Chip im Mehrzustandsfall.

#### Variable Steuerung der Spannung an das Steuer-Gate

Das Systemdiagramm von Fig. 5 im Zusammenhang mit den Tabellen 1 und 2 veranschaulicht, wie verschiedene Spannungen an die EEPROM-Anordnung 60 angelegt werden, um die Grundfunktionen des EEPROM zu erledigen. Bekannte EEPROM-Bausteine erlauben es der an das Steuer-Gate gelieferten Spannung  $V_{CG}$  nur, eine von zwei Spannungen, nämlich  $V_{CC}$  oder die höhere Programmiervoltage von etwa 12 V anzunehmen.

Bei einem weiteren Ausführungsbeispiel der vorliegenden Erfindung erlaubt es die an das Steuer-Gate angelegte Spannung  $V_{CG}$  über einen breiten Bereich an Spannungen unabhängig und kontinuierlich veränderlich zu sein. Dies wird ermöglicht durch  $V_{PG}$  vom Controller 140. In einzelnen wird  $V_{CG}$  auf einer Leitung 83 von  $V_{PG}$  gespeist, was wiederum vom Controller von einer Leitung 901 geliefert wird. Tabelle 2 zeigt, wie  $V_{PG}$  unter unterschiedlichen Funktionen des EEPROM verschiedene Spannungen annimmt.

Die Veränderlichkeit von  $V_{CG}$  ist besonders vorteilhaft bei Programmier- und Löschrandausnutzungssystemen. Bei der Programmierendausnutzung wird das Lesen während der Programmierverifizierung mit  $V_{CG}$  auf etwas höherer Spannung als der normalen  $V_{CC}$  vorgenommen. Das hilft dazu, die programmierte Schwelle gut in den Zustand hinein zu plazieren, indem an dem Trennschwellenpegel mit einer geringfügigen Marge vorbeiprogrammiert wird. Bei der Löschrandausnutzung wird die Zelle mit einer geringfügig reduzierten  $V_{CG}$  verifiziert, um die Zelle gut in den "gelöschten" Zustand zu bringen. Ferner kann die Randausnutzung angewandt werden, um das weiter oben beschriebene Problem der Beibehaltung der Ladung aufzuwiegen (Fig. 8B).

Wie schon erwähnt, arbeiten bekannte EEPROMs üblicherweise mit  $V_{CC}$ , um  $V_{CG}$  während der Programmier- oder Löschrandausnutzung zu speisen. Um eine Randausnutzung vorzunehmen, muß  $V_{CC}$  selbst stufenweise angehoben oder reduziert werden. Eine solche Praxis bringt ungenaue Ergebnisse in den Leseschaltungen hervor, da diesen auch  $V_{CC}$  als Antrieb dient.

Die Veränderlichkeit von  $V_{CG}$  unabhängig von der Leseschaltung zugeführten Spannungen erzeugen genauere und zuverlässigere Ergebnisse.

- 5 Ferner ist der breite Bereich von  $V_{CG}$  nützlich während des Prüfens und der Diagnose des EEPROM. Er erlaubt ein leichtes Messen der vollen Spanne der Schwelle der programmierten Zelle durch fortgesetztes Erhöhen von  $V_{CG}$  (bis zu dem Maximum, welches vom Übergangszusammenbruch des Bausteins begrenzt ist).
- 10 Zwar sind die hier beschriebenen Ausführungsbeispiele dieser Erfindung die bevorzugten Verwirklichungen, aber dem Fachmann ist klar, daß auch Abwandlungen derselben möglich sein können. Deshalb steht der Erfindung der Schutz innerhalb des vollen Umfangs der beigefügten Ansprüche zu.

EP 0 778 582

**PATENTANSPRÜCHE**

1. Elektrisch löschbares und programmierbares, nichtflüchtiges Speichersystem, umfassend:

eine integrierte Schaltungsanordnung elektrisch änderbarer Speicherzellen (11, 13);  
eine Programmierschaltung (210), die geeignete Programmierungsparameter parallel an eine adressierte Mehrzahl von Zellen anlegt,  
eine Einrichtung (200) zum parallelen Verifizieren des Zustands, in den die adressierte Mehrzahl von Zellen programmiert ist,  
eine Einrichtung zur Unterbindung einer weiteren Programmierung richtig verifizierter Zellen unter der Mehrzahl adressierter Zellen, und  
eine Einrichtung zum weiteren Programmieren und Verifizieren in paralleler Weise der Mehrzahl adressierter Zellen und zum Unterbinden einer Programmierung richtig verifizierter Zellen bis alle der Mehrzahl adressierter Zellen richtig verifiziert sind.

2. Speichersystem nach Anspruch 1, das weiterhin wenigstens eine Referenzspeicherzelle (400) enthält und das zusätzlich eine Einrichtung (410) zum Programmieren der wenigstens einen Referenzzelle auf einen Referenzwert ( $V_{T1}$ ) aufweist, und bei dem die Verifiziereinrichtung eine Einrichtung zum Lesen des Referenzwerts der wenigstens einen Referenzzelle zum Verifizieren der programmierten Daten enthält.

3. Speichersystem nach Anspruch 1, bei dem die Anordnung elektrisch änderbarer Speicherzellen (11, 13) in einer Mehrzahl adressierbarer Blöcke organisiert ist, die dadurch gekennzeichnet sind, daß die Zellen eines einzelnen Blocks zusammen löscherbar sind, wobei wenigstens eine Referenzzelle (400) in einzelnen der Blöcke von Zellen enthalten ist und das Speichersystem zusätzlich eine Einrichtung (411, 413) zum Programmieren der wenigstens einen Referenzzelle auf einen Referenzwert enthält, und bei dem die Verifiziereinrichtung eine Einrichtung zum Lesen des wenigstens einen Referenzwerts der Referenzzelle des Blocks enthält, wenn die darin enthaltenen programmierten Daten verifiziert werden.

4. Speichersystem nach Anspruch 1, bei dem die Einrichtung zum Unterbinden einer weiteren Programmierung richtig verifizierter Zellen eine Mehrzahl von Latch-Gliedern enthält sowie eine Einrichtung zum Setzen einzelner der Latch-Glieder (721) als Antwort auf entsprechende der Mehrzahl adressierter Zellen, die verifiziert werden.

5. Speichersystem nach Anspruch 1, bei dem die Einrichtung zum Verifizieren der programmierten Daten auf dem Chip eine Einrichtung zur Erfassung eines Parameters, der zu den Ladungswerten der einzelnen programmierten Zellen in Beziehung steht und, eine Einrichtung zum

Vergleichen des erfaßten Parameters von den einzelnen programmierten Zellen mit wenigstens einem Referenzparameter aufweist, der zu entsprechenden einzelnen Bits der Menge an Daten, die programmiert werden, in Beziehung steht, wobei einzelne programmierte Zellen verifiziert werden, sobald der Parametervergleich vollzogen ist.

5

6. Speichersystem nach Anspruch 1, ferner umfassend:

einen Controller (140) zur Steuerung des Betriebs der Speicherzellen (11, 13); und

eine Einrichtung zur Ausgabe eines Signals von dem Chip an den Controller zur Anzeige, daß die Mehrzahl adressierter Zellen verifiziert ist.

10

7. Speichersystem nach irgendeinem der Ansprüche 1 bis 6, bei dem die Speicherzellen (11, 13) einzeln mehr als zwei spezielle Zustände aufweisen.

8. Speichersystem nach irgendeinem der Ansprüche 1 bis 6, bei dem die Speicherzellen

15

(11, 13) einzeln genau zwei spezielle Zustände aufweisen.

9. Elektrisch löschbares und programmierbares, nichtflüchtiges Speichersystem, umfassend:

eine integrierte Schaltungsanordnung elektrisch änderbarer Speicherzellen (11, 13);

20

eine Löschschaltung, die geeignete Löschmodi parallel an eine adressierte Mehrzahl von Zellen anlegt;

eine Einrichtung (700) zum parallelen Verifizieren des Zustands, in dem die adressierte Mehrzahl von Zellen gelöscht ist,

25

eine Einrichtung zum Unterbinden eines weiteren Löschs richtig verifizierter Zellen unter der Mehrzahl adressierter Zellen, und

eine Einrichtung zum weiteren Löschen und Verifizieren in paralleler Weise der Mehrzahl adressierter Zellen und zum Unterbinden des Löschs richtig verifizierter Zellen bis alle der Mehrzahl adressierter Zellen richtig verifiziert sind.

30

10. Speichersystem nach Anspruch 9, das wenigstens eine Referenzspeicherzelle (400) enthält und zusätzlich eine Einrichtung zum Programmieren der wenigstens einen Referenzzelle auf einen Referenzwert ( $V_{T1}$ ) aufweist, und bei dem die Verifiziereinrichtung eine Einrichtung zum Lesen des Referenzwerts der wenigstens einen Referenzzelle zum Verifizieren des gelöschten Zustands enthält.

35

11. Speichersystem nach Anspruch 9, bei dem die Anordnung elektrisch änderbarer Speicherzellen (11, 13) in einer Mehrzahl adressierbarer Blöcke organisiert ist, die dadurch gekennzeichnet sind, daß die Zellen eines einzelnen Blocks zusammen löscher sind, wobei wenigstens eine Referenzzelle (400) in einzelnen der Blöcke von Zellen enthalten ist, und das Speichersystem zusätzlich eine Einrichtung zum Programmieren der wenigstens einen Referenzzelle auf einen Referenzwert enthält, und wobei die Verifiziereinrichtung eine Einrichtung zum Lesen des wenigstens einen Referenzwerts der Referenzzelle des Blocks enthält, wenn der Löscherzustand der darin befindlichen Zellen verifiziert wird.

40

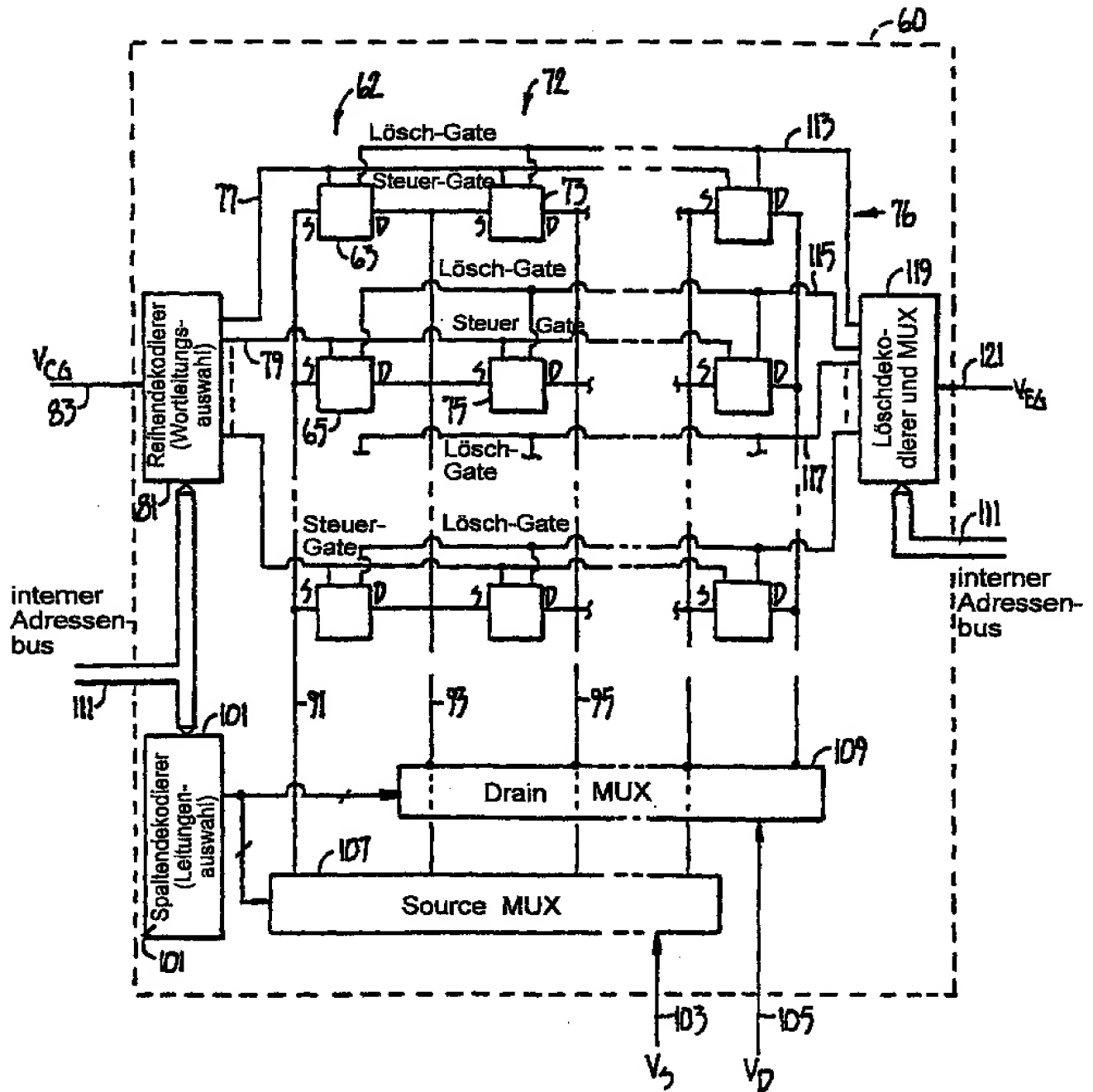
12. Speichersystem nach Anspruch 9, bei dem die Einrichtung zum Verifizieren des gelöschten Zustands auf dem Chip eine Einrichtung zum Erfassen eines Parameters, der mit den Ladungswerten der einzelnen gelöschten Zellen in Beziehung steht, und eine Einrichtung zum Vergleich des von den einzelnen gelöschten Zellen erfaßten Parameters mit wenigstens einem Referenzparameter enthält, welcher mit entsprechenden einzelnen Bits der Mehrzahl adressierter Zellen, die gelöscht werden, in Beziehung steht, wobei einzelne gelöschte Zellen verifiziert werden, sobald der Parametervergleich vollzogen ist.

13. Speichersystem nach Anspruch 9, ferner umfassend eine Einrichtung zum Programmieren der Zellen in dem Löschzustand zu dem Speicherzustand neben dem Löschzustand, um dadurch eine Gleichförmigkeit des Schwellenwerts in jeder der gelöschten Zellen sowie sicherzustellen, daß alle Zellen einem ähnlichen Grad an Programm/Löschbeanspruchung ausgesetzt werden.

14. Speichersystem nach irgendeinem der Ansprüche 9 bis 13, bei dem die Speicherzellen (11, 13) einzeln mehr als zwei spezielle Zustände aufweisen.

15. Speichersystem nach irgendeinem der Ansprüche 9 bis 13, bei dem die Speicherzellen einzeln genau zwei spezielle Zustände aufweisen.





**Fig. 4**



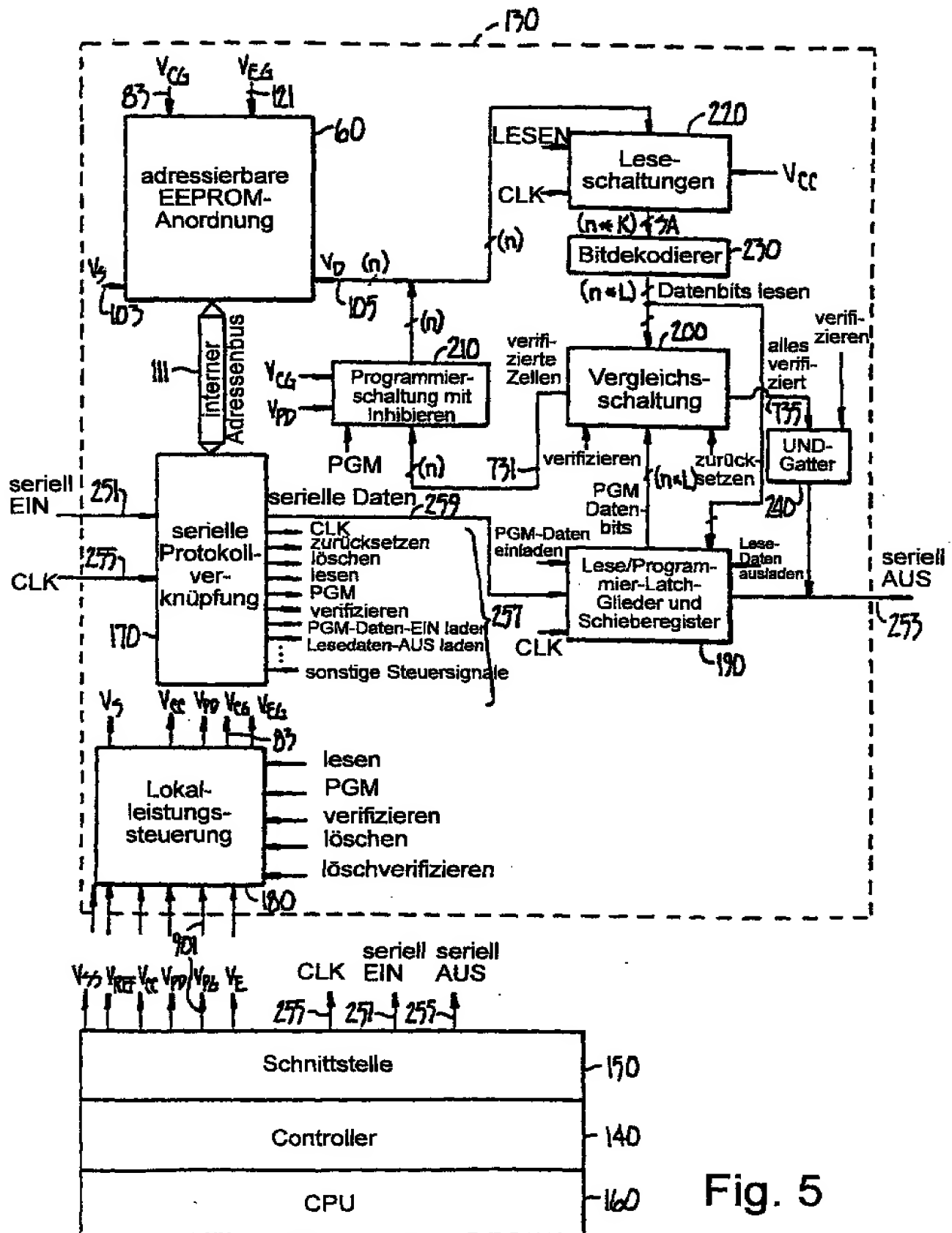


Fig. 5

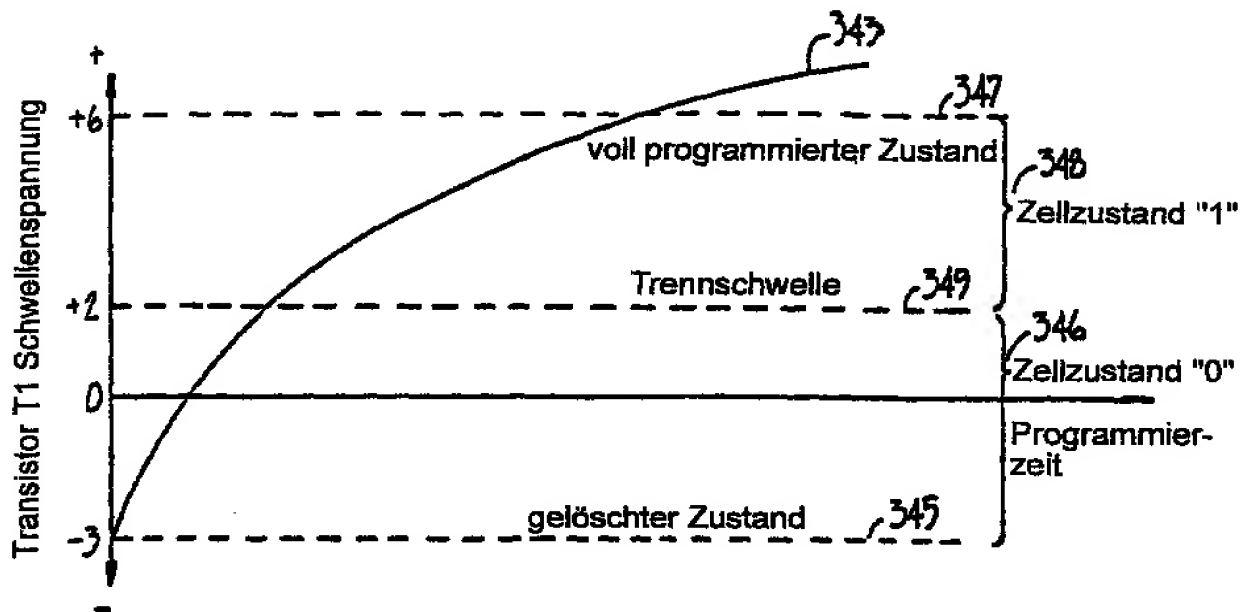


Fig. 6

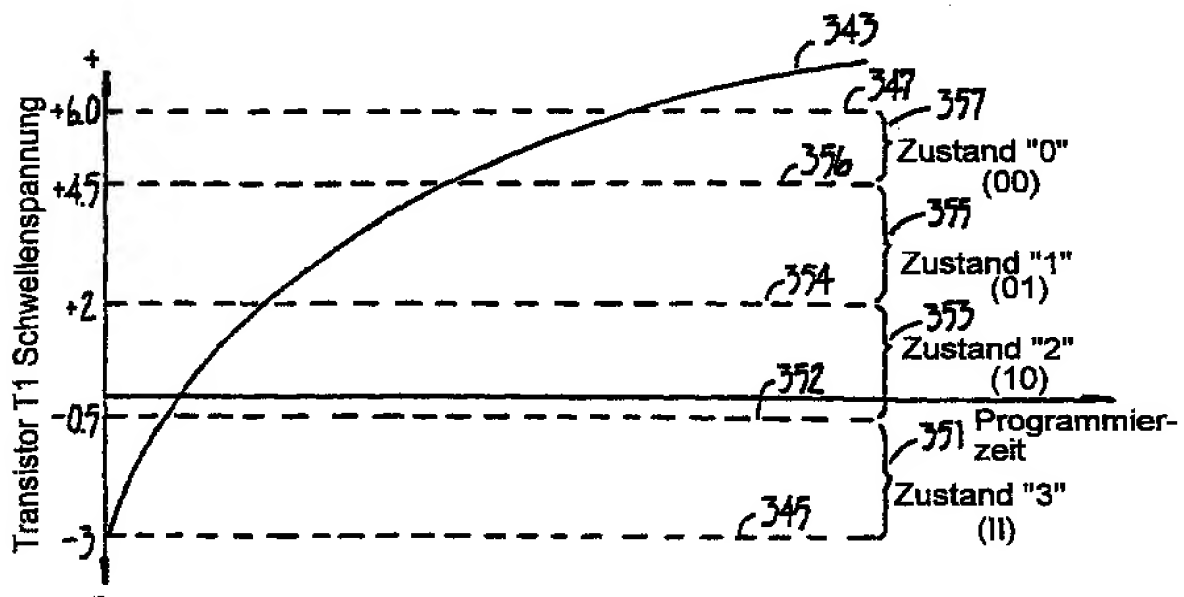


Fig. 7A

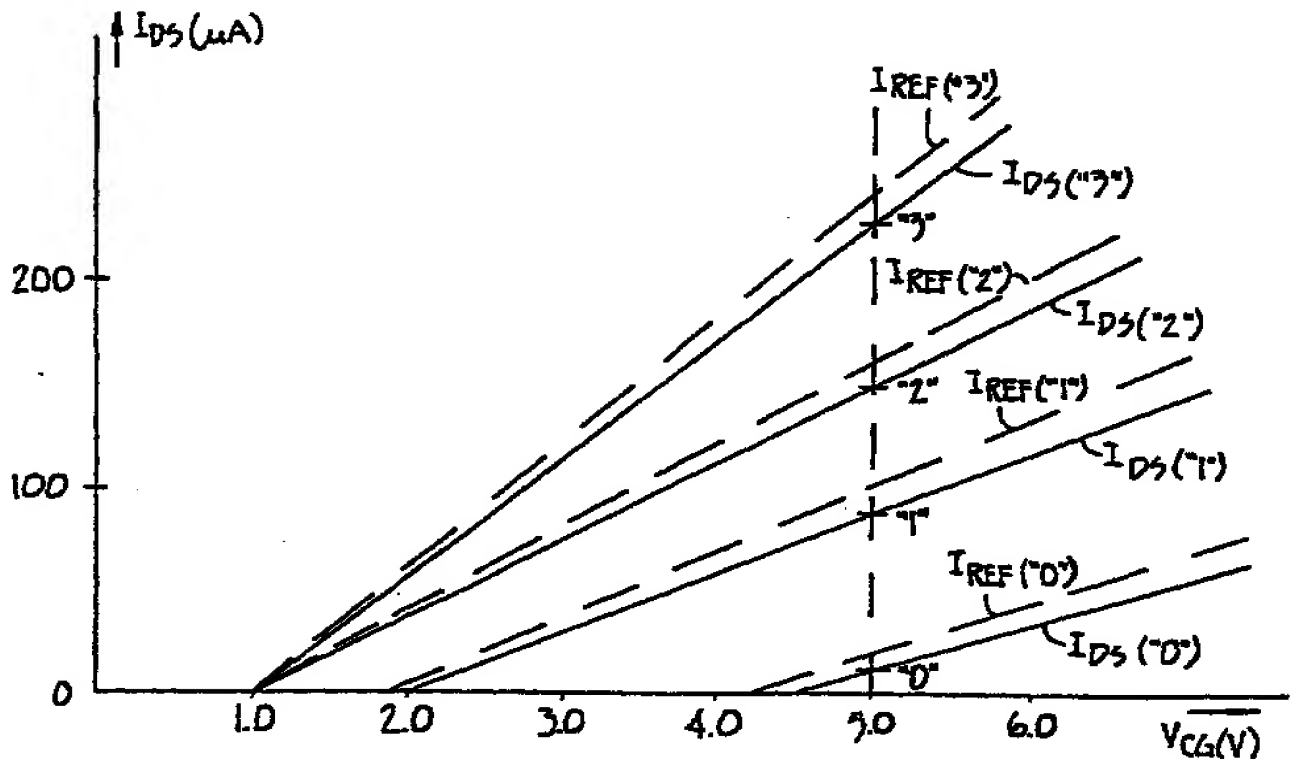


Fig. 7B

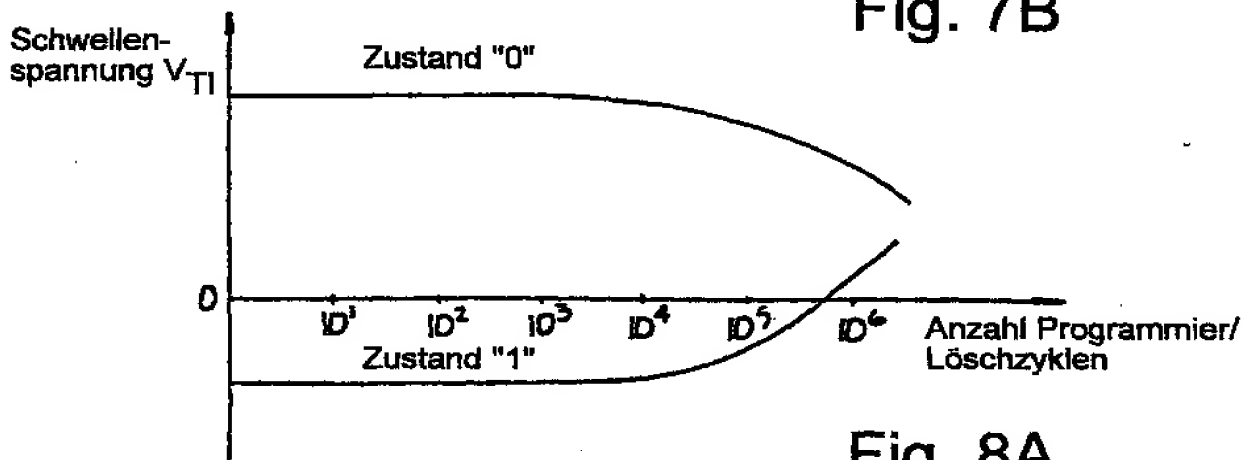


Fig. 8A

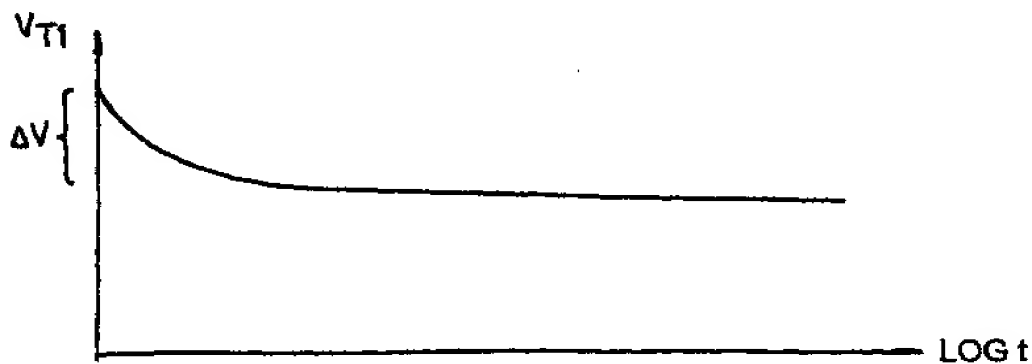


Fig. 8B

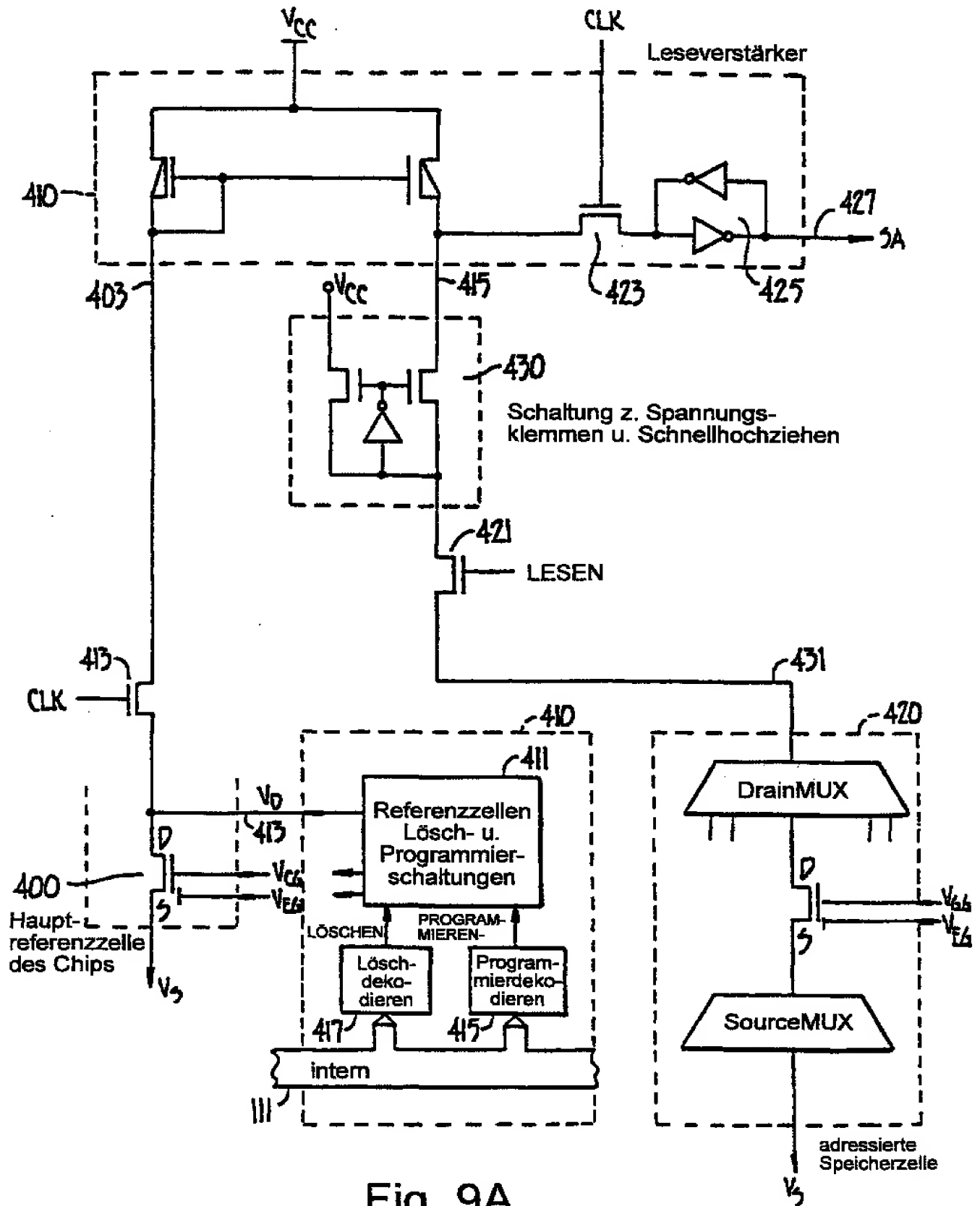


Fig. 9A

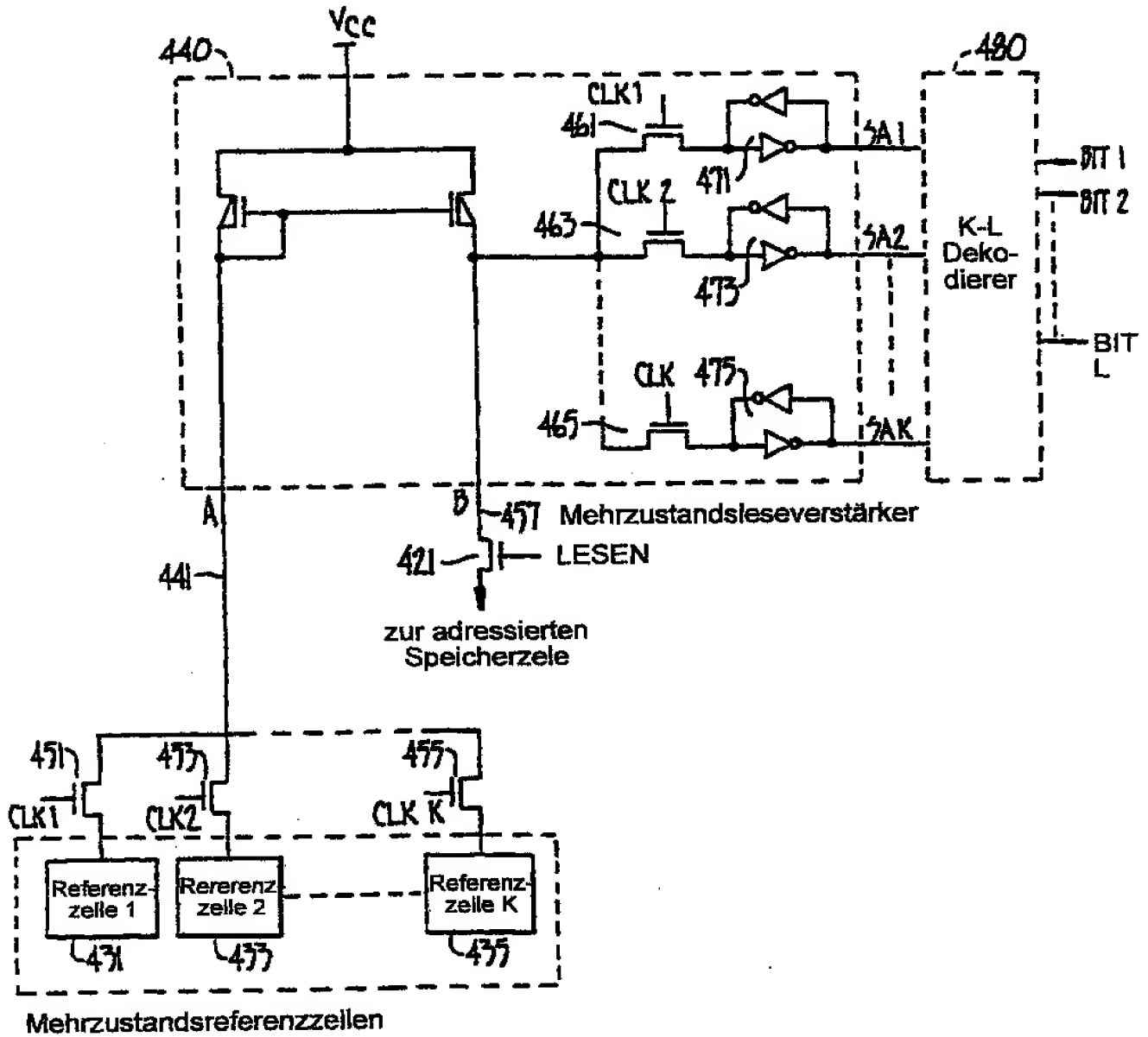


Fig. 9B

8/21

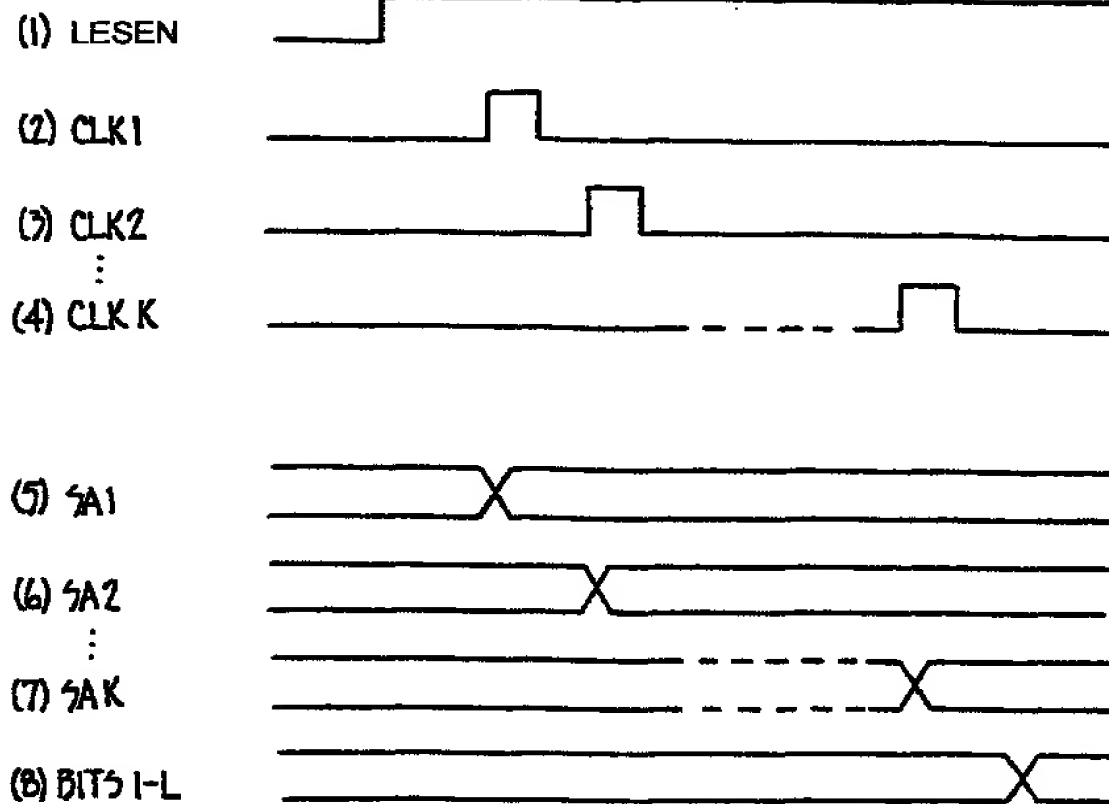


Fig. 9C

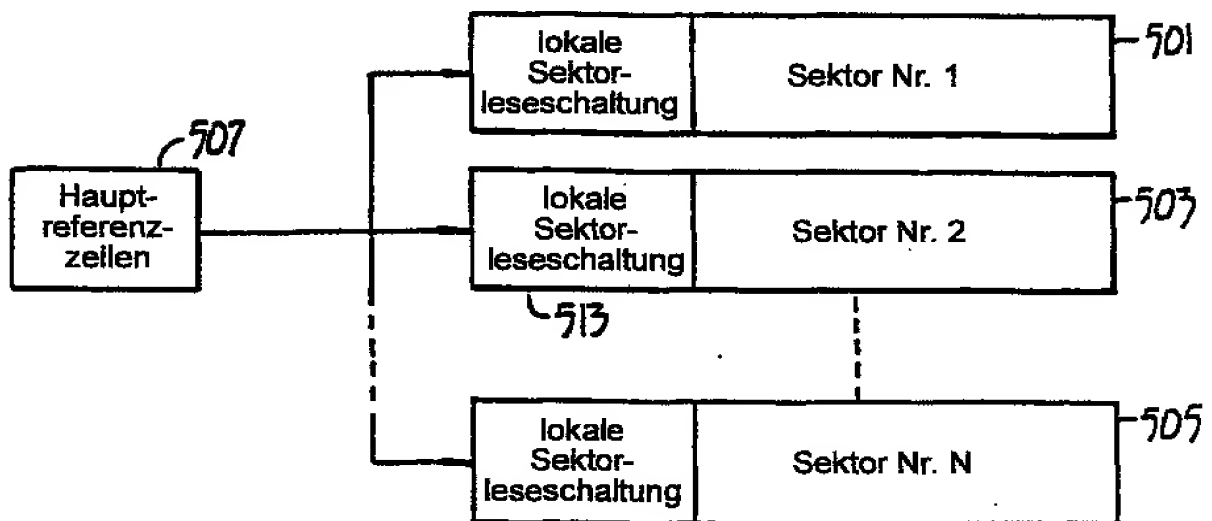


Fig. 10

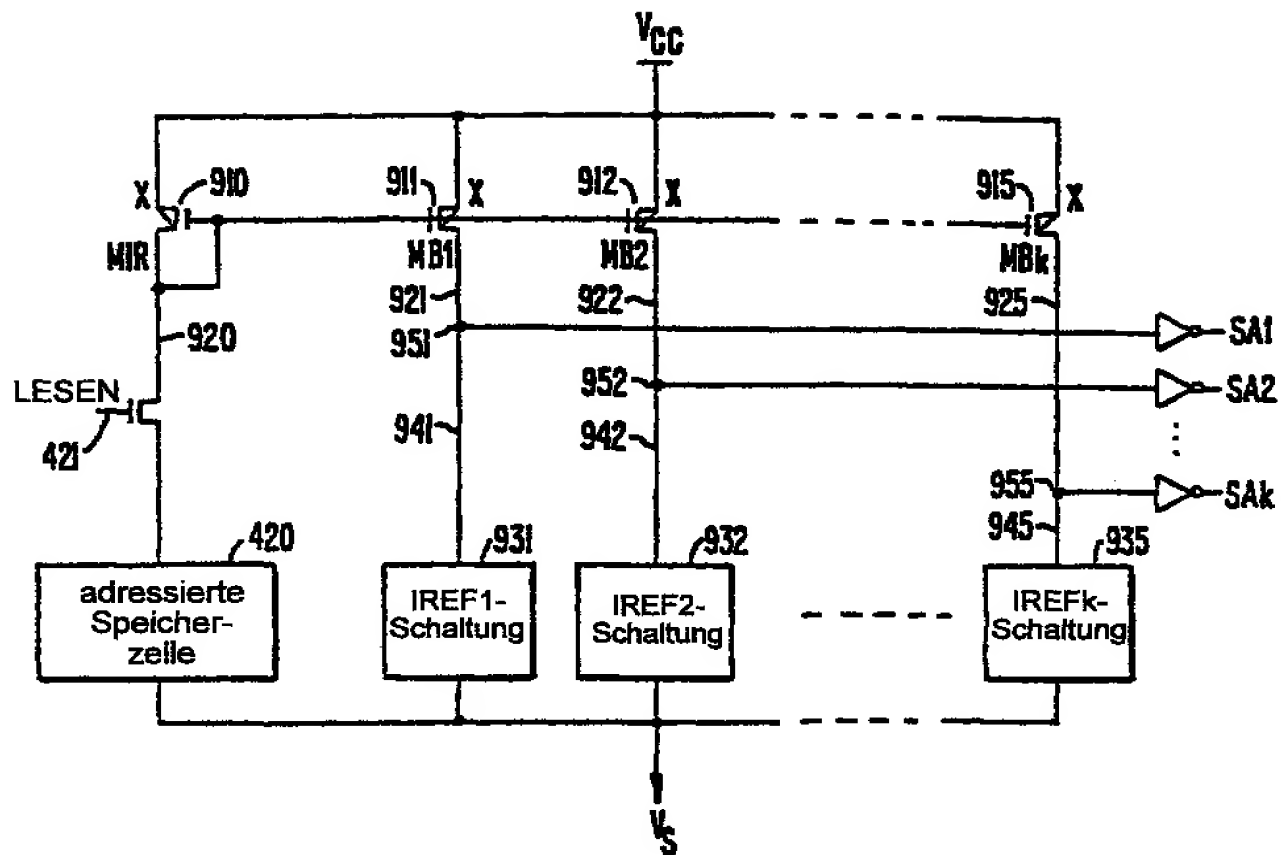


Fig. 9D

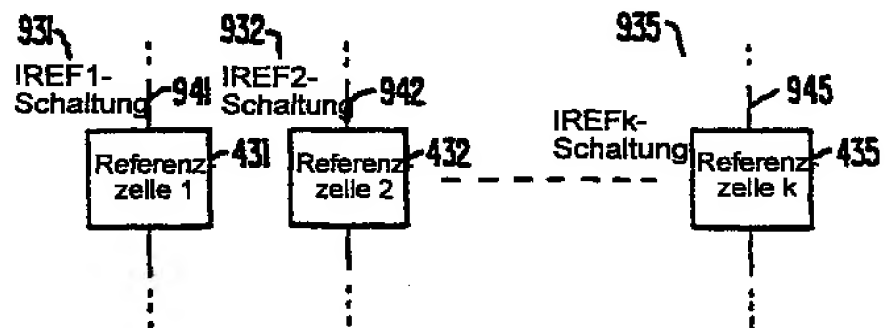


Fig. 9E

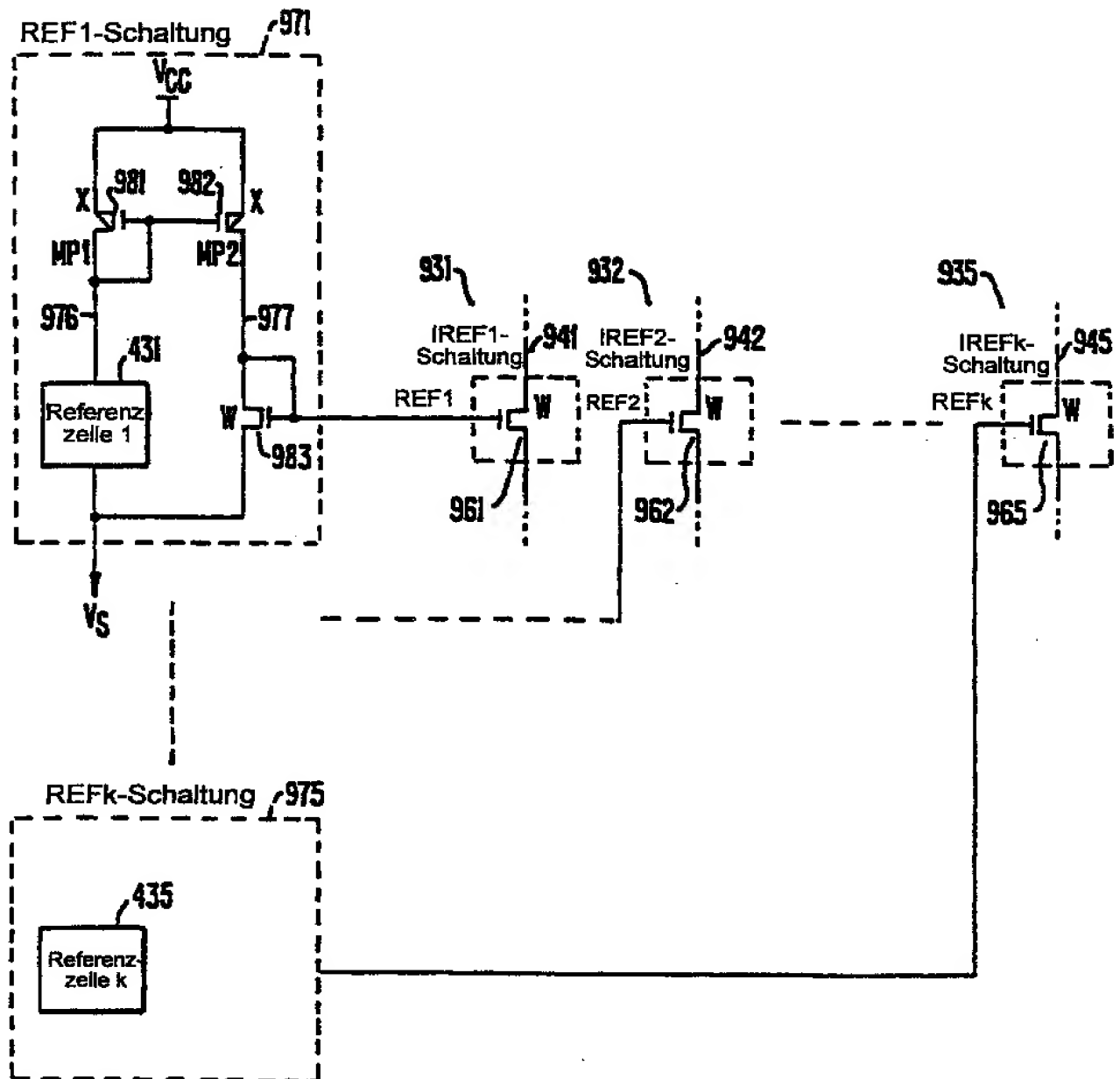


Fig. 9F



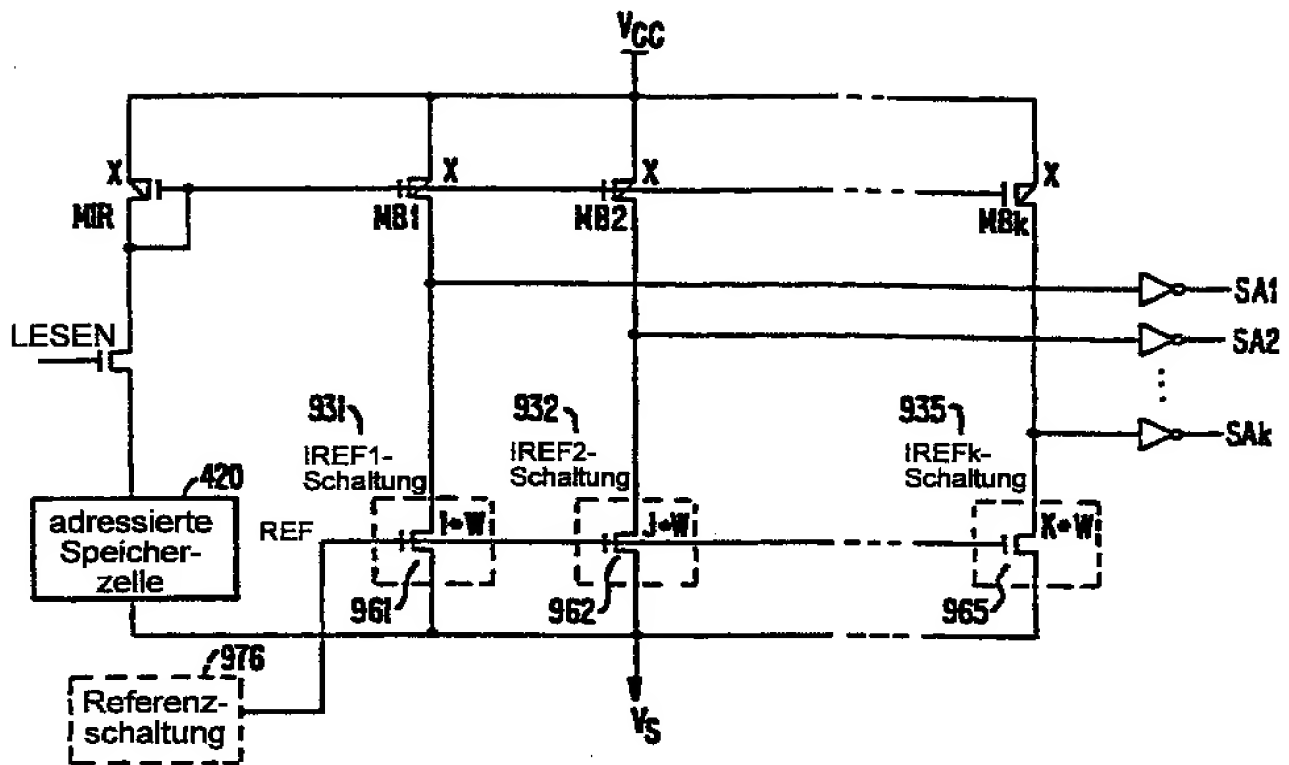


Fig. 9G

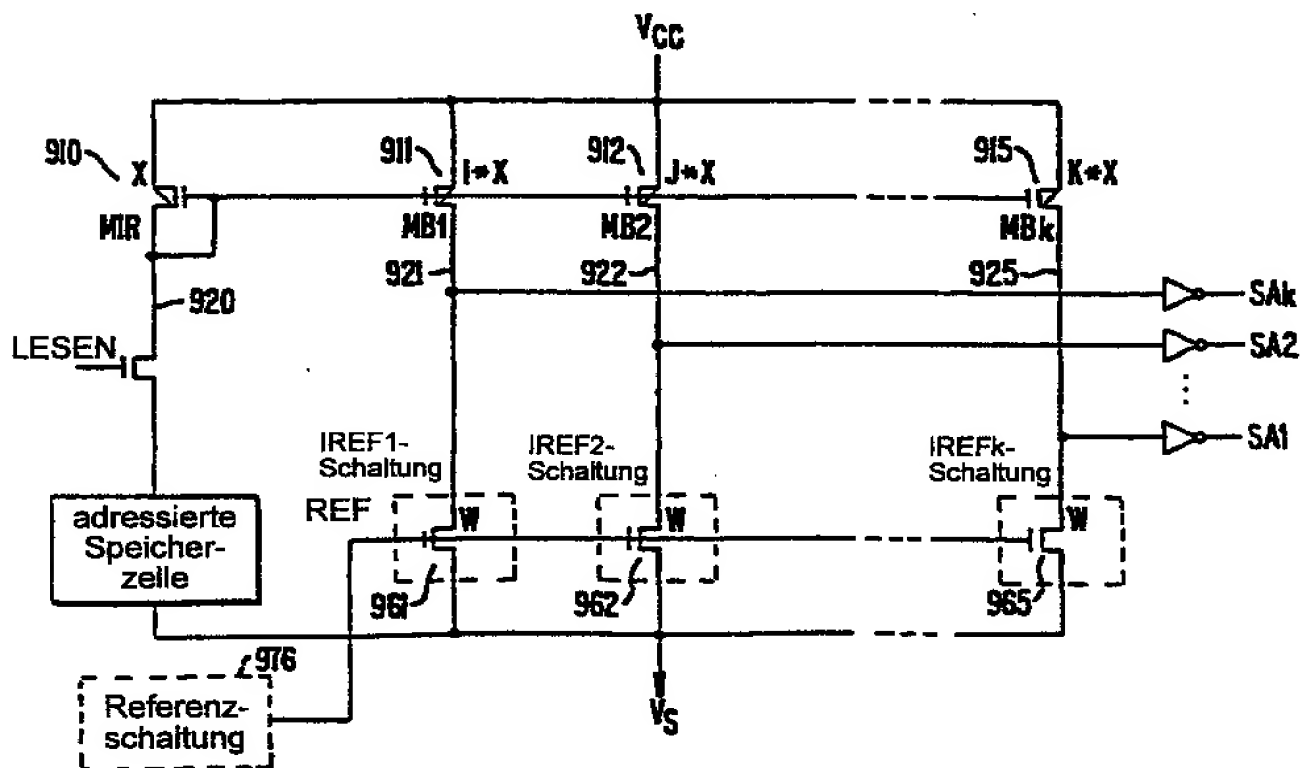


Fig. 9H

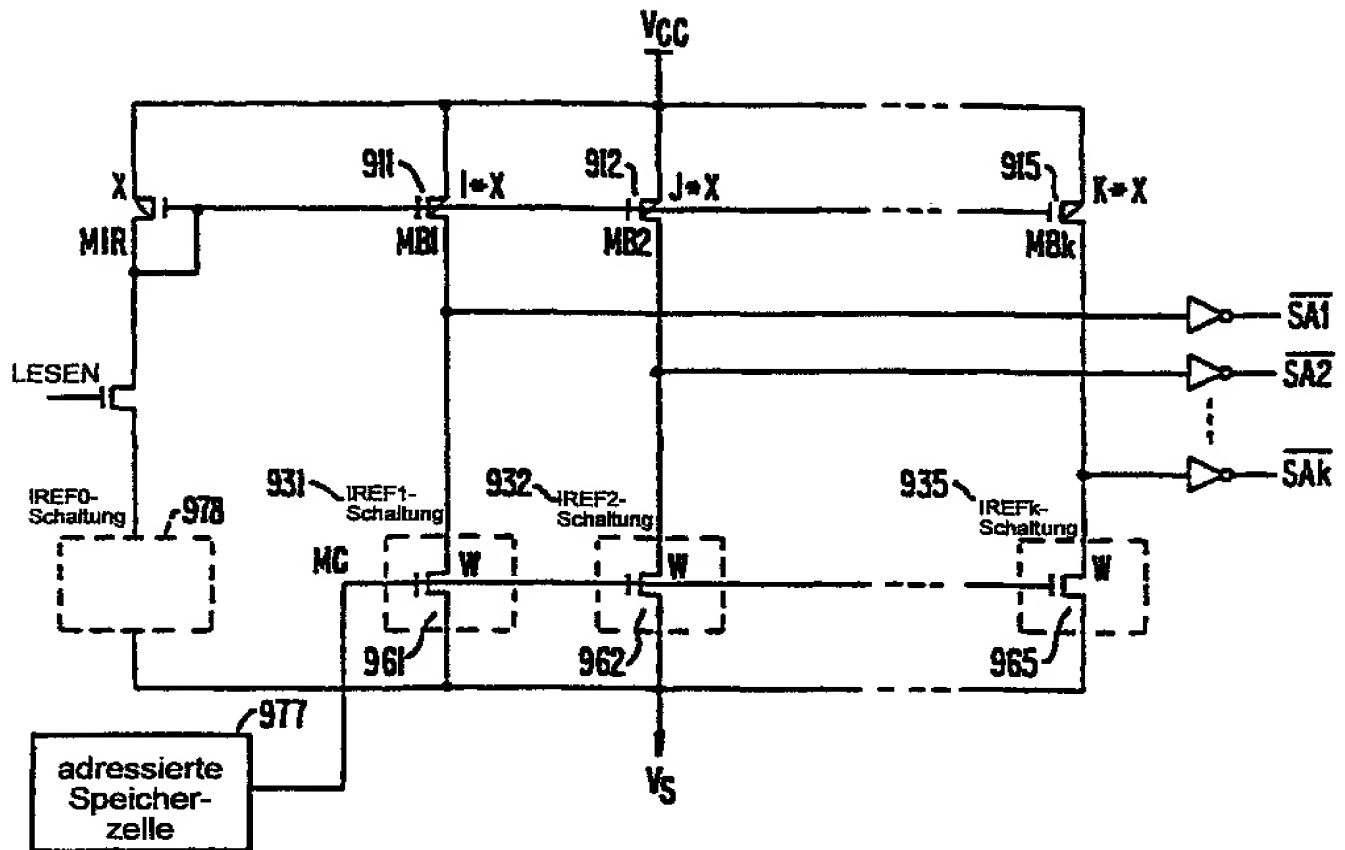
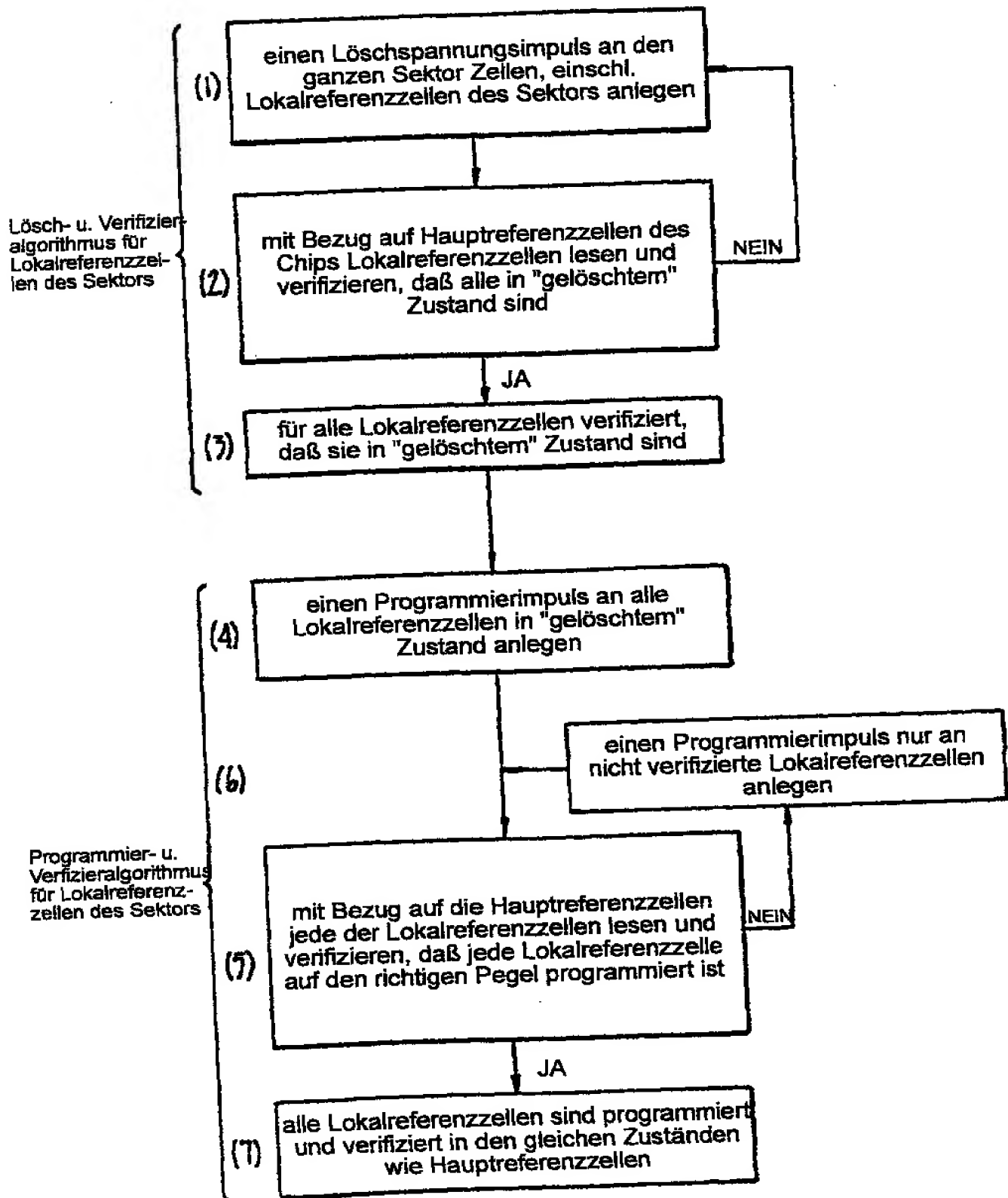


Fig. 9I



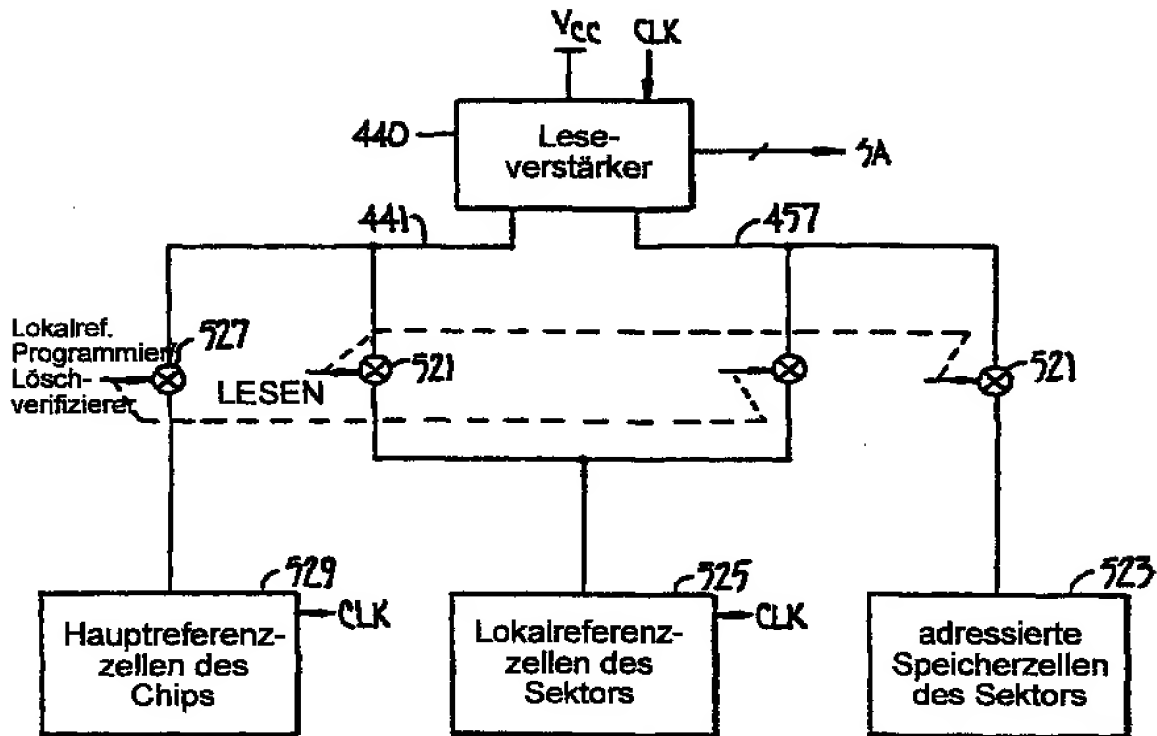


Fig. 12A

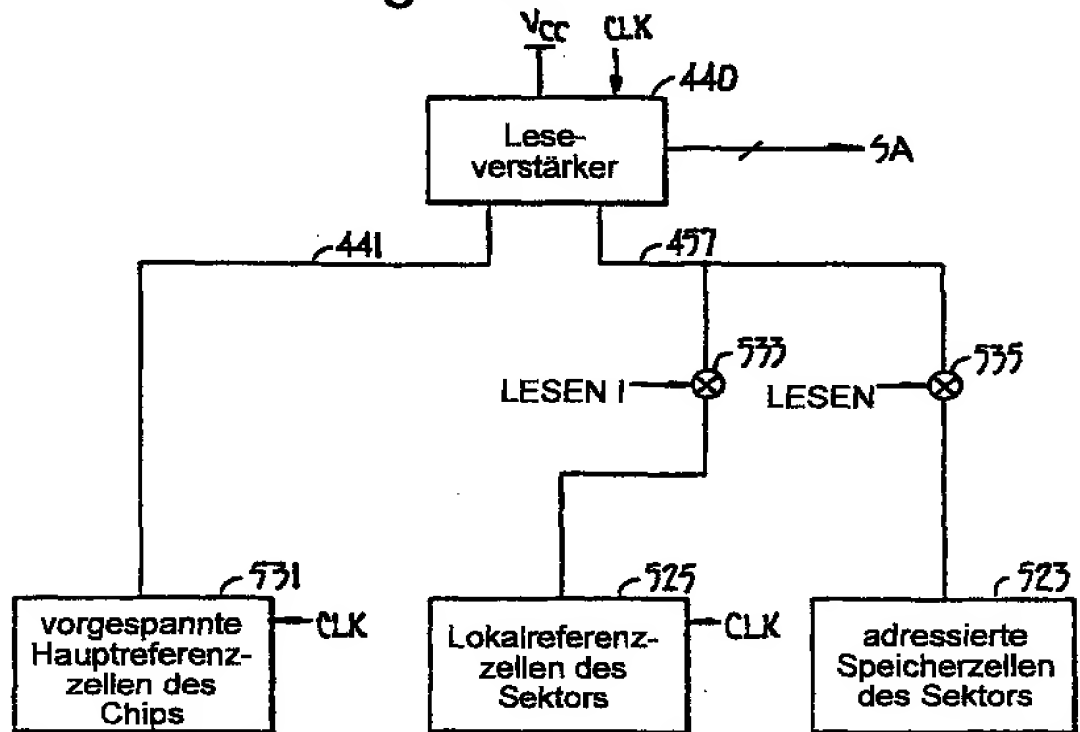


Fig. 13A

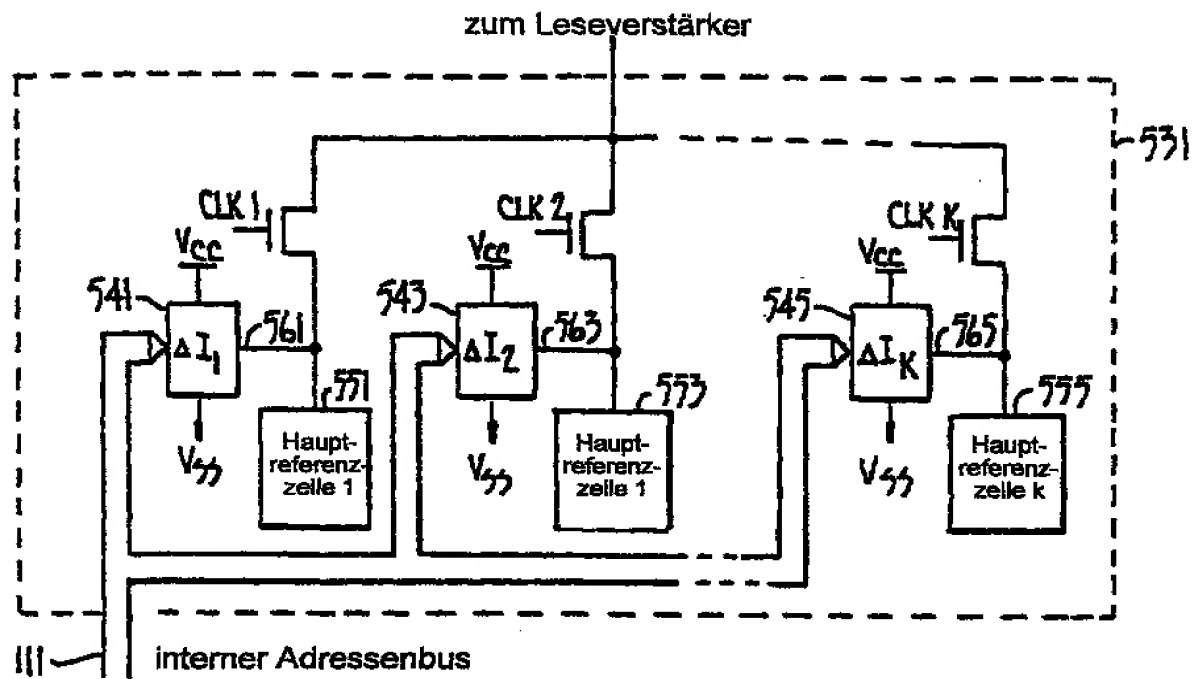


Fig. 13B

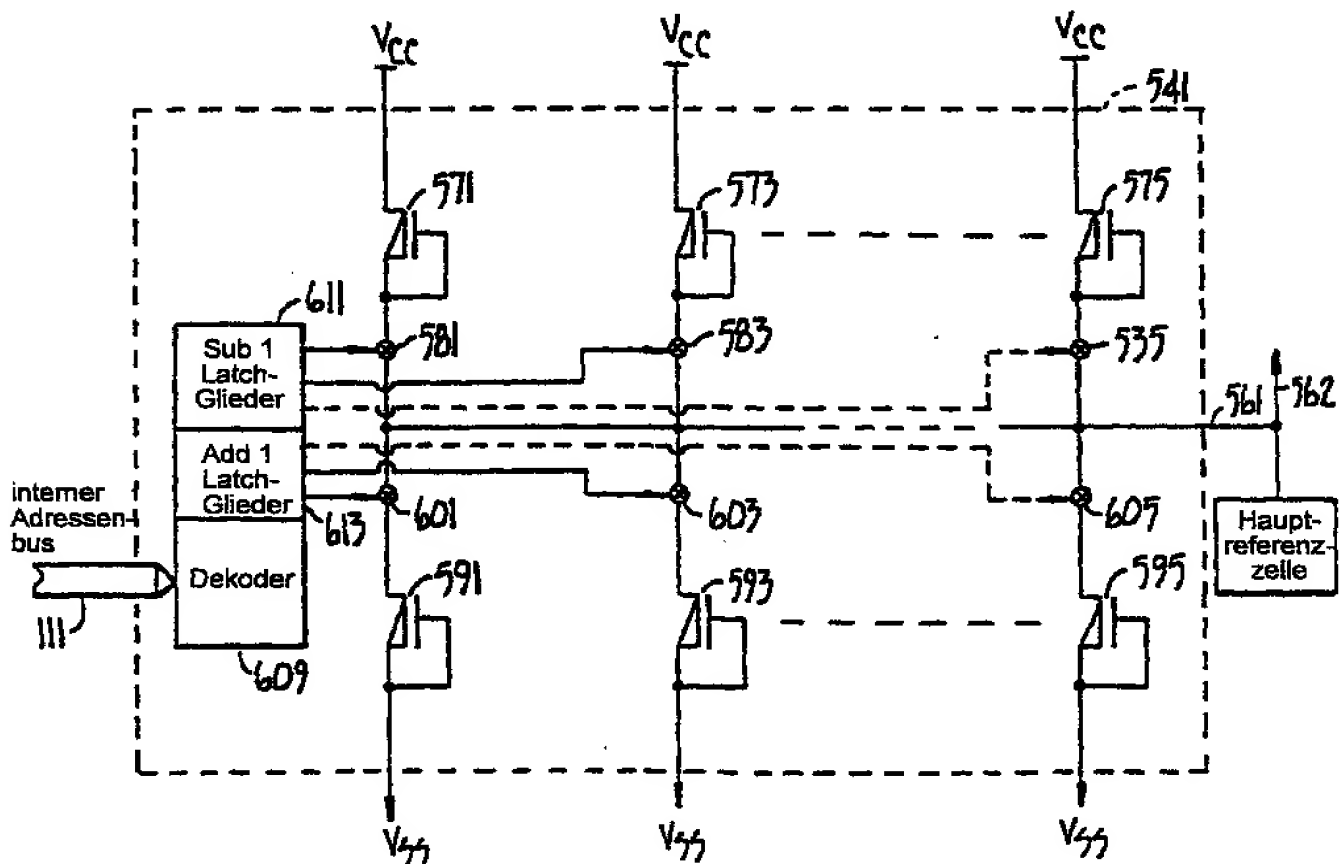


Fig. 13C

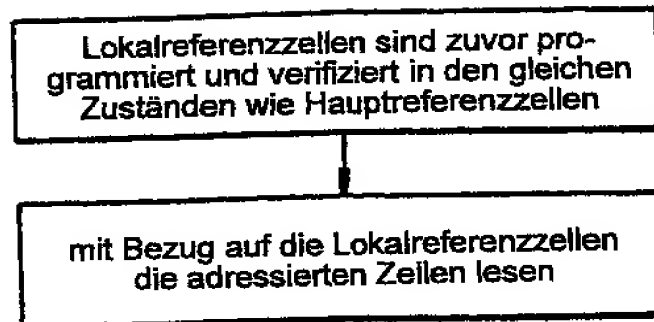


Fig. 12B

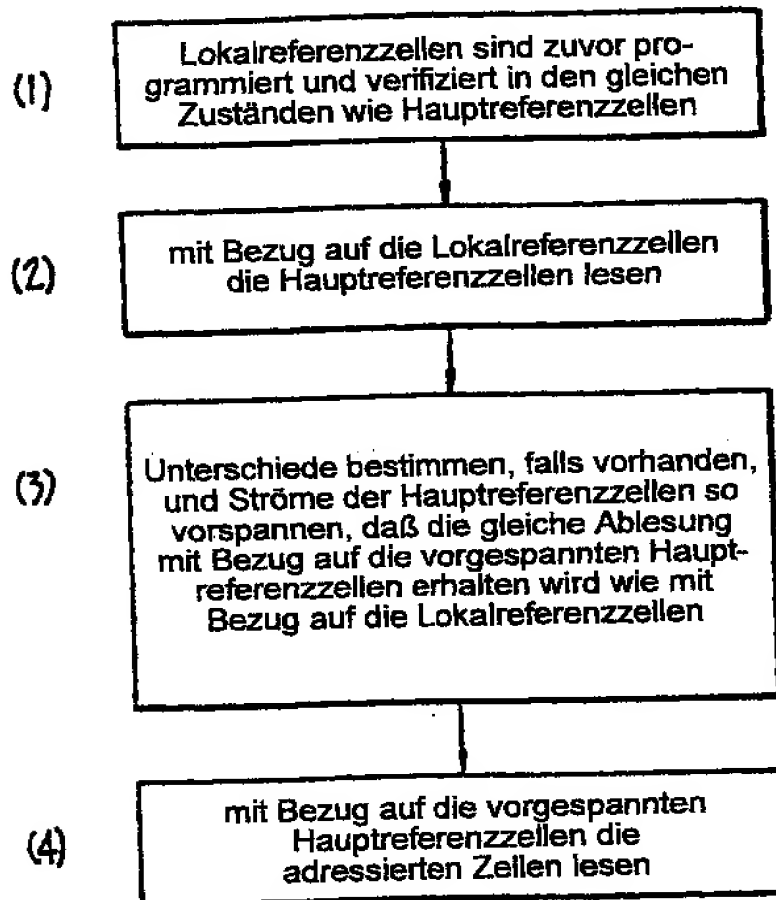
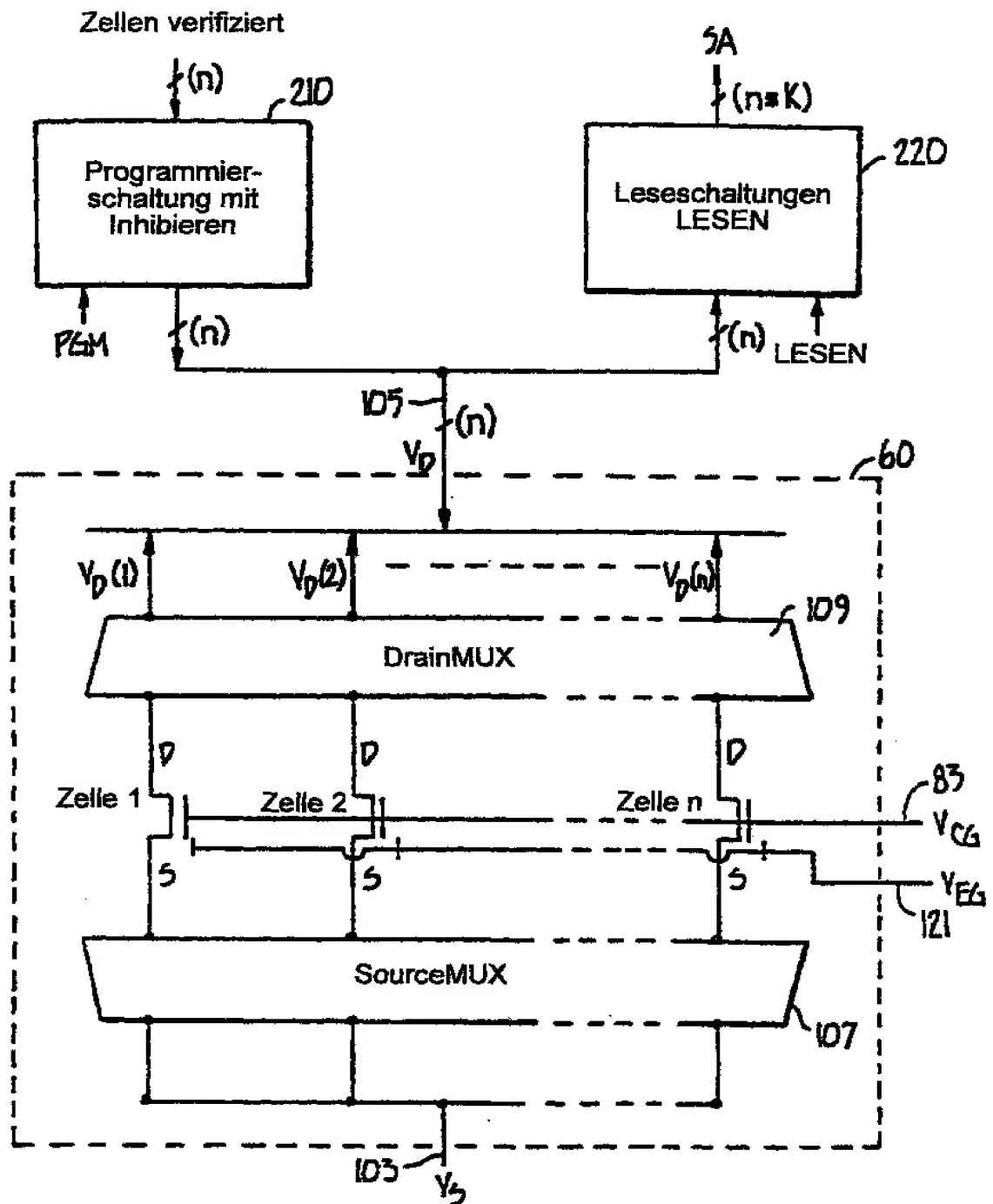


Fig. 13D



Lese/Programmierdatenpfade für  $n$  parallele Zellen

Fig. 14

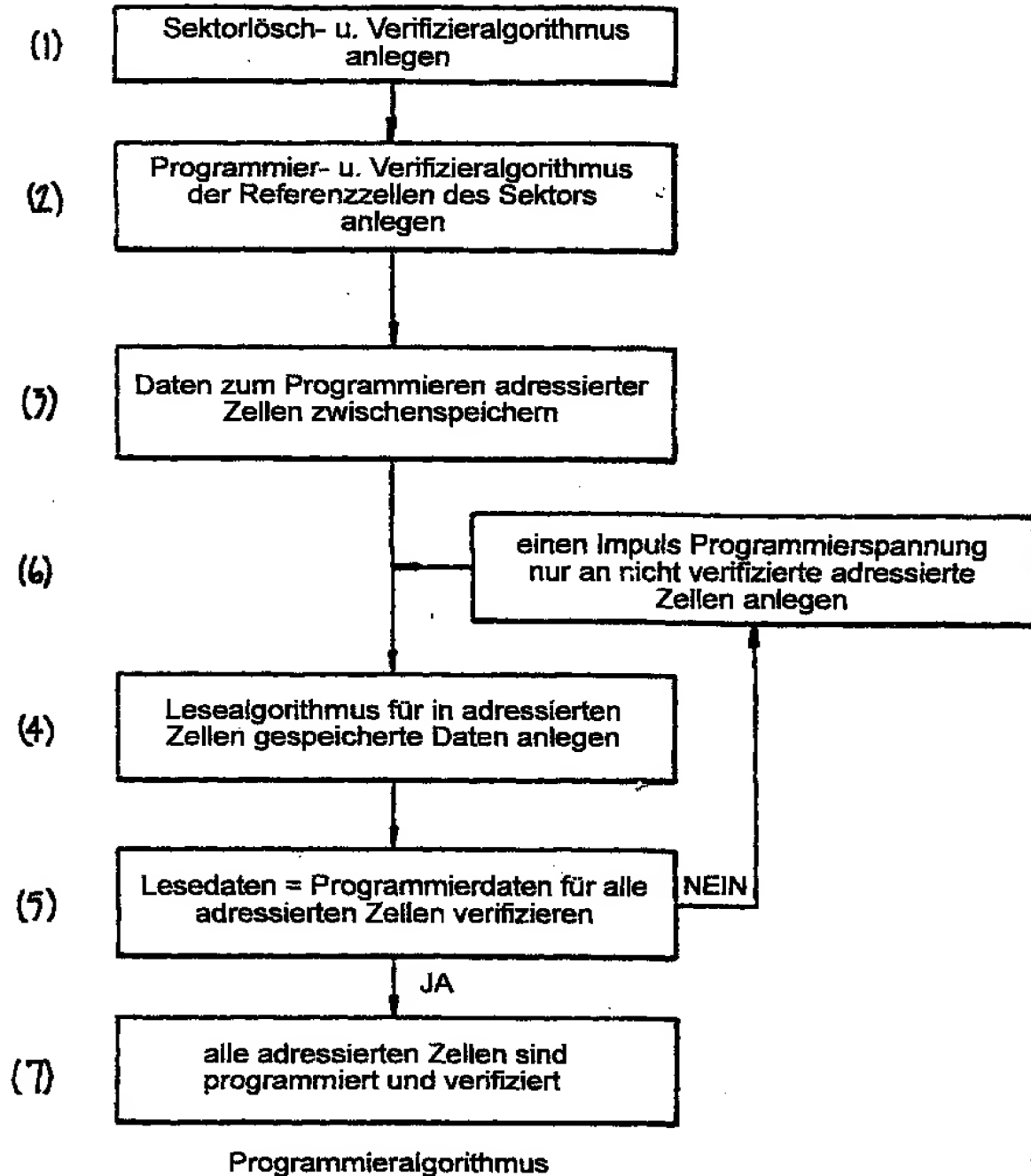


Fig. 15





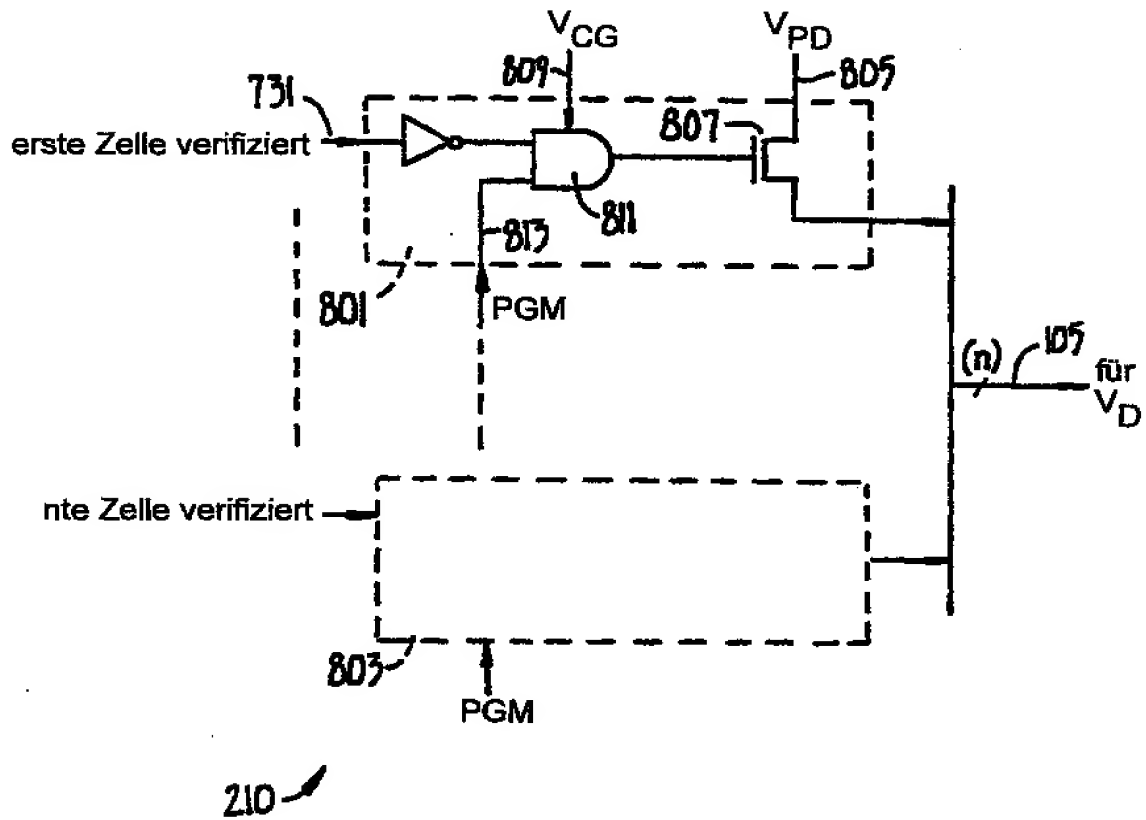


Fig. 17

	ausgewähltes Steuer-Gate $V_{CG}$	Drain $V_D$	Source $V_S$	Lösch- Gate $V_{EG}$
lesen	$V_{PG}$	$V_{REF}$	$V_{SS}$	$V_E$
program- mieren	$V_{PG}$	$V_{PD}$	$V_{SS}$	$V_E$
program- mierverifi- zieren	$V_{PG}$	$V_{REF}$	$V_{SS}$	$V_E$
löschen	$V_{PG}$	$V_{REF}$	$V_{SS}$	$V_E$
löschverifi- zieren	$V_{PG}$	$V_{REF}$	$V_{SS}$	$V_E$

Tabelle 1

(typische Werte)	lesen	program- mieren	program- mierverifi- zieren	löschen	lösch- verifizieren
$V_{PG}$	$V_{CC}$	12V	$V_{CC} + \delta V$	$V_{CC}$	$V_{CC} - \delta V$
$V_{CC}$	5V	5V	5V	5V	5V
$V_{PD}$	$V_{SS}$	8V	8V	$V_{SS}$	$V_{SS}$
$V_E$	$V_{SS}$	$V_{SS}$	$V_{SS}$	20V	$V_{SS}$
nicht-ausgewähltes Steuer-Gate	$V_{SS}$	$V_{SS}$	$V_{SS}$	$V_{SS}$	$V_{SS}$
nicht-ausgewählte Bitzeile	$V_{REF}$	$V_{REF}$	$V_{REF}$	$V_{REF}$	$V_{REF}$

$V_{SS} = 0V$ ,  $V_{REF} = 1.5V$ ,  $\delta V = 0.5V - 1V$

Tabelle 2